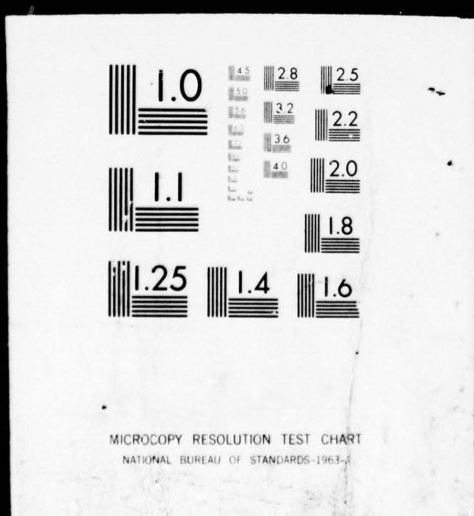
AD-A051 753 NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS F/G 12/1
SOME STATISTICAL PROCEDURES BASED ON DISTANCES.(U)
NOV 76 J J WALKER DAAG29-74-C-0030
UNCLASSIFIED MMS-1096 ARO-11959.19-M NL

1 OF 2
AD
A051 753

1 OF 2

AD
A051753



THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

SOME STATISTICAL PROCEDURES BASED ON DISTANCES

by

Joseph J. Walker

Institute of Statistics Mimeo Series #1096

November, 1976

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

JOSEPH J. WALKER. Some Statistical Procedures Based on Distances*
(Under the direction of I.M. CHAKRAVARTI and N.L. JOHNSON.)

A criterion is proposed for classifying multivariate "observations" according to their populations of origin when the observable data are the distances between pairs of "observations," with these distances themselves subject to further variation, such as measurement error. The same basic problem is investigated under several assumptions on the underlying normal distributions. In each case, the criterion is shown to be a particular quadratic form in normal variables. In the simplest case considered, a computational form for the distribution is given. An asymptotic expansion is developed which provides an approximation to the distribution in other cases. The accuracy of the approximation is investigated numerically.

The related problem of estimation of the noncentrality parameter of a noncentral chi-squared random variable is also investigated. An estimator is proposed which is based on the two-sample Wilcoxon statistic, using independent samples from the central and noncentral chi-squared distributions. The estimator has the property that it is invariant under monotonic transformations of the observed data. Further properties of the estimator are derived and its asymptotic relative efficiency with respect to the maximum likelihood estimator is investigated numerically.

---

*This research was supported in part by the Army Research Office (Contract DAAG29-74-C-0030) and the National Science Foundation (Grant CP 42325).

ACKNOWLEDGEMENTS

I wish to express my greatest thanks to Professors N.L. Johnson and I.M. Chakravarti, who served as advisors in this research. Their guidance and encouragement during the course of this investigation were invaluable.

My sincere thanks also go to the other members of the examining committee, Professors G.D. Simons, R.J. Carroll, and C.M. Suchindran. Their agreement to read the manuscript form of this dissertation was of great help, as were their comments on it, particularly those of Professor Simons.

I also wish to express my gratitude to all of the other members of the faculty of the Department of Statistics who have contributed to my education. In particular, I thank Professor E.J. Wegman, who was my advisor for some of my graduate years, and Professor M.R. Leadbetter, without whose help in obtaining a master's degree prior to being drafted, I probably would have been sent to Vietnam.

I am indebted to the U.S. Army Research Office (contract DAAG29-74-C-0030) and to the National Science Foundation (grant CP42325) for their partial support of the research contained in this dissertation. I am also grateful for the financial help received through the Veterans Administration.

For her excellent typing of the manuscript, I thank Ms. June Maxwell.

Finally, I must express my deep appreciation of the contributions of my parents. Their sacrifices made my undergraduate education possible, and their continued moral support and occasional emergency financial help sustained me throughout this endeavor.

TABLE OF CONTENTS

Preceding Page Blank - Filmed

1. INTRODUCTION

In most statistical problems which are analyzed through multivariate methods, it is assumed that we have several characteristics of interest and measurements of each of those characteristics for each of several individuals. Thus if there are p characteristics and n individuals, the data would consist of n p-dimensional vectors $\underline{x}_i' = (x_{i1},\ldots,x_{ip})$, i=1,...,n. Inferences concerning the distribution of the random vectors can then be made using the given data.

The primary aim of this investigation is to explore the making of inferences when the observable data are not the vectors described above, but rather some measurements of how "far apart" pairs of individuals are. Thus if we think of the n vectors as representing the positions of n points in p-dimensional space, each point representing an individual or object, then the observable data might be the euclidean (or some other) distance between pairs of points or individuals. To complicate the problem further, there could be additional variation arising out of errors of measurement. If our real interest is in making inferences concerning the underlying distribution of the vectors, then these additional measurement errors could be thought of as distortions of the "true" distances, those distances between the points with positions given by the vectors.

In some cases the underlying distribution may be masked even more. For example, the observed data may consist of some subjective judgments

or perceptions as to the degree of similarity (or dissimilarity) of the various individuals or objects. In such cases, we can still assume that there is some underlying distribution of characteristics of the individuals and that the perceptions of similarity are related in some way to the distances between individuals in the characteristic space. Shepard [25] gives a summary of data of this type and various other related types along with a discussion of several methods of analysis.

Before discussing the specific topics of this investigation, we shall briefly mention several examples involving data of this sort. In these cases, methods of analysis have been developed for making inferences based on distances or distance-like data.

Paired comparison analysis is a technique for obtaining an ordering of a set of objects with respect to some property, which is usually measured only in a subjective way (e.g. taste). The procedure is discussed extensively by David [5]. Typically, the objects are presented in pairs to judges who indicate which of each pair rates higher. The total number of judges rating object i higher than object j, less the number rating j higher than i (in absolute value), could be interpreted as a measure of dissimilarity between the two objects: the closer to zero the dissimilarity is, the less sure the judges are about the difference (i.e., the "closer together" the objects are). Here the "measurement error" does not enter directly into the final measure of dissimilarity, but rather is a factor in the judges' decisions as to which objects rate higher.

Cluster analysis embraces a wide variety of techniques for dividing a collection of objects into groups, in such a way that those in a given

group are more similar to each other according to some criterion than they are to those in the other groups. There are a number of reviews of the various techniques in the literature (e.g. Cormack [3], Hartigan [10]). Although not all clustering methods use distance or similarity data, a great many do. The similarities are sometimes subjective measures or judgments as in paired comparison methods. In other cases, they are direct measurements or estimates of the distances between pairs of objects. In fact, even when the vectors of observed characteristics are available, those data are often converted to a similarity matrix, for example by calculating the correlation coefficients among the characteristics. Some of the results obtained in this investigation, while not applicable directly to the clustering problem, are related to it, and might be adapted to it as criteria for deciding when objects belong to the same cluster.

A problem which is related to clustering, but is considerably simpler to analyze statistically is that of classification or discrimination, specifically the classification of one or more objects as belonging to one of several known groups or populations. An even more basic problem is that of making a decision as to whether one or more objects belongs to or does not belong to a given group or to one of several groups. When the underlying distributions of the populations are normal, such problems have been extensively investigated (e.g. Anderson [2], chapter 6, Kendall and Stuart [15], chapter 44). Some work has also been done with nonparametric classification rules (e.g. Das Gupta [4]).

All of the classification techniques mentioned in the above references, however, assume that the various characteristics can themselves

be measured. The author is unaware of any similar results using only distance or dissimilarity data. Chapters two and three of this dissertation address those questions when the underlying population distributions are normal and the observed distances between objects contain additional measurement error. In Chapter two, a rule for deciding whether an object is from a certain population or not is proposed, and its distribution in each case is investigated. Chapter three contains extensions of this basic decision problem to more complicated situations: specifically, the inclusion of several objects to be classified and the problem of assigning an object to one of two known populations.

It will be seen that the results obtained here involve the distributions of various quadratic forms in normal variables. While most of the results used here are taken from Johnson and Kotz [11], chapter 29, it should be pointed out that they were originally derived by others. For example, we have used a representation due to Gurland [8],[9]. Other early work in the area was done, for example, by Robbins [22]. A representation due to Press [20], mentioned in Section 3.3, may allow generalization of some of the results obtained here.

Chapter four is not connected directly with the other two main chapters, but it is indirectly related to them. Under the assumption of normality for the distribution of the measurement error, the resulting distances are related to noncentral chi-squared random variables (with the noncentrality parameter indicating the true distance). A related problem is the estimation of the noncentrality parameter. If we have estimates of the distance between the means of two (possibly) different populations, we can use them to make a judgment as to whether the true

distance between population means is positive, i.e. whether the population are different. Maximum likelihood estimation of the noncentrality parameter leads to a rather complicated equation which often does not have a simple solution (see, e.g. Johnson and Kotz [11], p. 136). In Chapter four, we propose a simple estimator based on the two-sample Wilcoxon statistic and investigate some of its properties.

Finally, mention should be made here of the technique called multidimensional scaling, which was introduced by Shepard [23],[24] and Kruskal [16],[17]. The basic goal of the procedure is to obtain a representation of a set of objects as a relatively low dimensional configuration of points, such that the distances between pairs of points closely correspond to the observed distances or dissimilarities between the respective objects. The methods used are iterative, successively adjusting the positions of the points until the rank ordering of the distances is as similar as possible to that of the dissimilarities, according to some criterion. Much of the motivation for the research presented here came as a result of attempts to find a rigorous statistical analysis for the scaling problem. That goal was not achieved and to the author's knowledge, has not been achieved by others either. Perhaps some of the results given here will find application at a later date to more complicated scaling problems.

2. BASIC CLASSIFICATION

2.1 Introduction

As indicated in Chapter one, the classification of objects according to the populations from which they originated has been studied quite extensively. If we assume that each observation consists of the measurements of p characteristics for a given object, the assignment to a population can be made based on these measurements (see, e.g. Anderson [2], chapter 6). The classification criteria which result often can be interpreted in terms of the "distances" of the observations from the populations in question. For example, consider the problem of classifying a single p-variate observation as being from one of two known multivariate normal populations (see Anderson [2], section 6.4). If the observed measurements are $\underline{X}' = (X_1,\ldots,X_p)$ and the population distributions, are p-variate normal distributions, denoted $N_p(\underline{\mu}^{(i)},\underline{\Sigma})$, i=1,2, where $\underline{\mu}^{(i)'} = (\mu_1^{(i)'},\ldots,\mu_p^{(i)})$ and $\underline{\Sigma}$ is a p×p positive definite symmetric matrix, then classification based on the discriminant function

$$\underline{X}'\underline{\Sigma}^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$$

can be shown to be optimal in terms of expected loss. Addition of the constant

$$\tfrac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\underline{\Sigma}^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$$

to the discriminant function shows that the classification can be made equivalently on the basis of the difference

$$(\underline{X} - \underline{\mu}^{(1)})'\underline{\Sigma}^{-1}(\underline{X} - \underline{\mu}^{(1)}) - (\underline{X} - \underline{\mu}^{(2)})'\underline{\Sigma}^{-1}(\underline{X} - \underline{\mu}^{(2)}) \ ,$$

that is, on a comparison of the Mahalanobis distance of the observation from the centers (means) of the two populations.

In this chapter, and in the following chapter, we wish to take this dependence on distances a step further. We will still assume that a p-dimensional normal distribution exists for the populations under consideration and that the object to be classified has p characteristics. However, these variates cannot be observed directly. The only observations which can be made are the distances of the object to be classified from other objects which are known to be from the given populations. We will also assume that an additional error is made in the determination of each distance. As discussed in the previous chapter, this additional error might correspond to measurement error or to the uncertainty introduced in the similarity-dissimilarity types of measures. Specifically, in this chapter we will investigate the question of whether a given object is from a single given population or not, and in the next chapter we will investigate several extensions to more complicated situations. More complete descriptions of the models used will be given as they are needed.

2.2 A criterion for the one population problem.

Probably the most basic problem which can be analyzed under the models considered here is whether a single new individual or object is from a given population or not. Before investigating this problem under the distance model, let us consider it from the classical point of veiw. Suppose we have a p-variate observation $\underline{X}' = (X_1,\ldots,X_p)$ and wish to decide whether it has arisen from a normal distribution with mean $\underline{\mu}' = (\mu_1,\ldots,\mu_p)$ and covariance matrix $\underline{\Sigma}$. Assuming that both $\underline{\mu}$ and $\underline{\Sigma}$ are known, we could consider the statistic

$$s = (\underline{X} - \underline{\mu})'\underline{\Sigma}^{-1}(\underline{x} - \underline{\mu}) ,$$

that is, the Mahalanobis distance of the observation from the mean of the distribution. If $\underline{X}$ is from the given population, then s is distributed as a chi-squared random variable with p degrees of freedom, and we can use that distribution for making inferences concerning whether $\underline{X}$ is from the population.

Returning to the problem based only on distance observations, we can derive a comparable result based on the distances of the new object from others known to be from the population. First, however, we must specify more completely the distance model which we are utilizing.

(A) We assume that there are n individuals known to be from a population and one individual which may or may not be from the same population. Let $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ and $\underline{X}_0$ be the respective (unobservable) values of the p-dimensional random variable upon which the classification is to be based, where for $i = 0,1,\ldots,n$, $\underline{X}_i' = (X_{i1},\ldots,X_{ip})$. We will assume that the population distribution is normal with mean $\underline{\mu}$ and covariance matrix $\underline{\Sigma}$. Thus $\underline{X}_0, \underline{X}_1, \ldots, \underline{X}_n$ are independent random variables, with $\underline{X}_1,\ldots,\underline{X}_n$ each distributed as $N_p(\underline{\mu},\underline{\Sigma})$ and $\underline{X}_0$ distributed as $N_p(\underline{\mu}_0,\underline{\Sigma})$. We assume that the additional error and resulting distance measurements have the following structure: let $\underline{Y}_{01}, \underline{Y}_{10}, \underline{Y}_{02}, \underline{Y}_{20}, \ldots, \underline{Y}_{0n}, \underline{Y}_{n0}$ be independent, identically distributed normal random variables with mean $\underline{0}$ and covariance matrix $\underline{\Delta}$, and for $i = 1,2,\ldots,n$, let

$$S_{0i} = (\underline{X}_0 + \underline{Y}_{0i} - \underline{X}_i - \underline{Y}_{i0})'\underline{\Sigma}^{-1}(\underline{X}_0 + \underline{Y}_{0i} - \underline{X}_i - \underline{Y}_{i0}) .$$

Then $S_{0i},\ldots,S_{0n}$ are the observed distances of the new individual from the n individuals known to be from the population. Notice that we let $\underline{Y}_{ij}$ be the error made in determining the position of individual i in the measurement of the distance from i to j and we assume all such errors to be independent. However, $S_{01},\ldots,S_{0n}$ are not independent since all depend on the value of the random variable $\underline{X}_0$.

Since $\underline{X}_1,\ldots,\underline{X}_n$ are dispersed about their population mean, the natural analogue to the distance from the mean in the simpler case would be the average distance of the individual to be classified from those known to be from the population, i.e.,

$$\overline{S}_0 = n^{-1}\sum_{i=1}^{n} S_{0i} .$$

This is the criterion which we shall investigate. Let us first, however, examine the relationship between $\overline{S}_0$ and the criterion we would use if we could actually observe $\underline{X}_0,\underline{X}_1,\ldots,\underline{X}_n$, in order to judge the appropriateness of $\overline{S}_0$. We will use the following lemma:

Lemma 2.2.1. If $\underline{X}_0,\underline{X}_1,\ldots,\underline{X}_n$ are arbitrary p-dimensional vectors and $\underline{A}$ is a symmetric matrix, then

$$\sum_{i=1}^{n}(\underline{X}_0-\underline{X}_i)'\underline{A}(\underline{X}_0-\underline{X}_i)$$
$$= n\sum_{i=1}^{n}(\underline{X}_0-\overline{\underline{X}})'\underline{A}(\underline{X}_0-\overline{\underline{X}}) + \sum_{i=1}^{n}(\underline{X}_i-\overline{\underline{X}})'\underline{A}(\underline{X}_i-\overline{\underline{X}}) ,$$

where $\overline{\underline{X}} = n^{-1}\sum_{i=1}^{n}\underline{X}_i$.

Proof: Since we can write $\underline{A} = \underline{BB}'$ for $\underline{B}$ appropriately chosen and trace($\underline{BC}$) = trace($\underline{CB}$) if $\underline{B}$ and $\underline{C}$ are conformable, we have

$$\sum_{i=1}^{n} (\underline{X}_0-\underline{X}_i)'\underline{A}(\underline{X}_0-\underline{X}_i) = \mathrm{tr}\{\sum_{i=1}^{n} \underline{A}(\underline{X}_0-\underline{X}_i)(\underline{X}_0-\underline{X}_i)'\}$$

$$= \mathrm{tr}\{\underline{A}\sum_{i=1}^{n} (\underline{X}_0-\underline{X}_i)(\underline{X}_0-\underline{X}_i)'\}$$

$$= \mathrm{tr}\{\underline{A}\sum_{i=1}^{n} (\underline{X}_0-\overline{\underline{X}} + \overline{\underline{X}}-\underline{X}_i)(\underline{X}_0-\overline{\underline{X}} + \overline{\underline{X}}-\underline{X}_i)'\}$$

$$= \mathrm{tr}\{\underline{A}[\sum_{i=1}^{n} (\underline{X}_0-\overline{\underline{X}})(\underline{X}_0-\overline{\underline{X}})' + \sum_{i=1}^{n} (\overline{\underline{X}}-\underline{X}_i)(\overline{\underline{X}}-\underline{X}_i)']\}$$

$$= n(\underline{X}_0-\overline{\underline{X}})'\underline{A}(\underline{X}_0-\overline{\underline{X}}) + \sum_{i=1}^{n} (\underline{X}_i-\overline{\underline{X}})'\underline{A}(\underline{X}_i-\overline{\underline{X}}) \qquad \square$$

Returning to the question of whether $\underline{X}_0$ is from the same population as $\underline{X}_1,\ldots,\underline{X}_n$, we see that, if the $\underline{X}$'s were observable, we would have a simple test of hypothesis situation: we have $\underline{X}_0$ which is distributed $N_p(\underline{\mu}_0,\underline{\Sigma})$ and $\underline{X}_1,\ldots,\underline{X}_n$ which are $N_p(\underline{\mu},\underline{\Sigma})$, and we wish to test the hypothesis H_0: $\underline{\mu}_0 = \underline{\mu}$ versus the alternative H_1: $\mu_0 \neq \underline{\mu}$. This is a standard two-sample problem and the usual methods of test construction (e.g. likelihood ratio) would lead to rejection of H_0 if $Q(\underline{X}) > c$ where

$$Q(\underline{X}) = (\underline{X}_0-\overline{\underline{X}})'\underline{\Sigma}^{-1}(\underline{X}_0-\overline{\underline{X}}) .$$

We cannot observe the $\underline{X}$'s themselves or $Q(\underline{X})$ either, but $\overline{S}_0$, which is observable, is related to $Q(\underline{X})$. For the moment, let us ignore the errors made in determining the positions of the $\underline{X}$'s for the distance measurements. (For convenience here, we shall refer to these errors, the $\underline{Y}$'s in the model, as contaminating variables, since they distort the positions of the $\underline{X}$'s.) Then by Lemma 2.2.1 we would have

$$\bar{S}_0 = n^{-1}\sum_{i=1}^{n} S_{0i} = n^{-1}\sum_{i=1}^{n} (\underline{X}_0-\underline{X}_i)'\underline{\Sigma}^{-1}(\underline{X}_0-\underline{X}_i)$$

$$= (\underline{X}_0-\bar{\underline{X}})'\underline{\Sigma}^{-1}(\underline{X}_0-\bar{\underline{X}}) + n^{-1}\sum_{i=1}^{n} (\underline{X}_i-\bar{\underline{X}})'\underline{\Sigma}^{-1}(\underline{X}_i-\bar{\underline{X}}) .$$

The first quantity on the right-hand side of the above equation is just $Q(\underline{X})$, and the second quantity is independent of $\underline{X}_0$ and of $\bar{\underline{X}}$ and hence of $Q(X)$; it is non-negative and does not depend on the hypothesis being tested. Thus a rejection region consisting of large values of $\bar{S}_0$ and one consisting of large values of $Q(X)$ will be approximately equivalent in the sense that a given set of $\underline{X}$'s will tend to lead to rejection for either region or acceptance for either, regardless of which hypothesis is true. Thus $\bar{S}_0$ would seem to be a reasonable test criterion to use in this simplified case.

Reintroducing the contaminating variables, we can express $\bar{S}_0$ as

$$\bar{S}_0 = n^{-1}\sum_{i=1}^{n} (\underline{X}_0-\underline{X}_i)'\underline{\Sigma}^{-1}(\underline{X}_0-\underline{X}_i)$$

$$+ 2n^{-1}\sum_{i=1}^{n} (X_0-\underline{X}_i)'\underline{\Sigma}^{-1}(\underline{Y}_{0i}-\underline{Y}_{i0})$$

$$+ n^{-1}\sum_{i=1}^{n} (\underline{Y}_{0i}-\underline{Y}_{i0})'\underline{\Sigma}^{-1}(\underline{Y}_{0i}-\underline{Y}_{i0}) .$$

By Lemma 2.2.1 and algebraic manipulation we have

$$\bar{S}_0 = (\underline{X}_0-\bar{\underline{X}})'\underline{\Sigma}^{-1}(\underline{X}_0-\bar{\underline{X}} + \underline{Z})$$

$$+ n^{-1}\sum_{i=1}^{n} (\underline{X}_i-\bar{\underline{X}} - \underline{Y}_{0i}+\underline{Y}_{i0})'\underline{\Sigma}^{-1}(\underline{X}_i-\bar{\underline{X}} - \underline{Y}_{0i}+\underline{Y}_{i0}) ,$$

where $\underline{Z} = 2n^{-1}\sum_{i=1}^{n}(\underline{Y}_{0i}-\underline{Y}_{i0})$. As before, the second factor on the right-hand side does not depend on the hypothesis being tested. The first factor, however, differs slightly from $Q(\underline{X})$; there is some distortion due

to the contaminating variables. But $\underline{Z}$ is independent of $\underline{X}_0$ and $\overline{\underline{X}}$, has expected value $\underline{0}$ and covariance matrix $8n^{-1}\underline{\Delta}$. Thus if we assume that $\underline{\Delta}$ is small relative to $\underline{\Sigma}$, the first factor differs only slightly from $Q(\underline{X})$, and thus a rejection region based on large values of $\overline{S}_0$ would still appear to be a reasonable one to use.

Following the proofs of two lemmas, we shall investigate the null distribution of $\overline{S}_0$, that is the distribution when $\underline{\mu}_0 = \underline{\mu}$.

Notation: In the following lemma and thereafter where needed, $\underline{I}_n$ will denote the n×n identity matrix, $\underline{J}_n$ will denote the n×n matrix consisting entirely of 1's, $\underline{0}$ will denote a matrix consisting entirely of 0's and $\otimes$ will denote the Kronecker product of two matrices: if $\underline{A} = ((a_{ij}))$ is p×p and $\underline{B}$ is q×q, then $\underline{A} \otimes \underline{B}$ is the pq×pq matrix

$$\begin{bmatrix} a_{11}\underline{B} & a_{12}\underline{B} & \cdots & a_{1p}\underline{B} \\ a_{21}\underline{B} & a_{22}\underline{B} & \cdots & a_{2p}\underline{B} \\ \vdots & & & \\ a_{p1}\underline{B} & a_{p2}\underline{B} & \cdots & a_{pp}\underline{B} \end{bmatrix} .$$

Lemma 2.2.2. Let $\underline{A}$ and $\underline{B}$ be p×p matrices and let $\underline{D} = \underline{I}_n \otimes \underline{A} + \underline{J}_n \otimes \underline{B}$. Then

$$|\underline{D}| = |\underline{A}|^{n-1}|\underline{A}+n\underline{B}| .$$

(Note: This lemma is a variation of Exercise 1.3 in Rao [21].)

Proof: For $k = 1,2,\ldots,p$ and $j = 1,2,\ldots,n-1$, add the $(k+jp)$-th column of the matrix to the k-th column, giving

$$|\underline{D}| = \begin{bmatrix} \underline{A}+n\underline{B} & \underline{B} & \underline{B} & \cdots & \underline{B} \\ \underline{A}+n\underline{B} & \underline{A}+\underline{B} & \underline{B} & \cdots & \underline{B} \\ \underline{A}+n\underline{B} & \underline{B} & \underline{A}+\underline{B} & \cdots & \underline{B} \\ \vdots & \vdots & \vdots & & \vdots \\ \underline{A}+n\underline{B} & \underline{B} & \underline{B} & \cdots & \underline{A}+\underline{B} \end{bmatrix} .$$

For $k = 1,2,\ldots,p$ and $j = 1,2,\ldots,n-1$, subtract the k-th row from the (k+jp)-th row, giving

$$|\underline{D}| = \begin{bmatrix} \underline{A}+n\underline{B} & \underline{B} & \underline{B} & \cdots & \underline{B} \\ \underline{0} & \underline{A} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{0} & \underline{A} & \cdots & \underline{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{A} \end{bmatrix} .$$

But if $\underline{C}$ and $\underline{E}$ are square matrices, then

$$\begin{bmatrix} \underline{C} & \underline{F} \\ \underline{0} & \underline{E} \end{bmatrix} = |\underline{C}||\underline{E}| .$$

Applying this fact, first with $\underline{C} = \underline{A}+n\underline{B}$ and then successively with $\underline{C} = \underline{A}$, we obtain the result. □

Lemma 2.2.3. Let $\underline{C}$ be a symmetric $p\times p$ matrix having characteristic roots $\lambda_1,\ldots,\lambda_p$ and matrix of associated orthogonal vectors $\underline{R}$. Let $\underline{A} = \underline{I}_n \otimes \underline{C} + \underline{J}_n \otimes \underline{I}_p$. Denote the characteristic roots of $\underline{A}$ by $\alpha_1,\alpha_2,\ldots,\alpha_{np}$. Then for $j = 1,2,\ldots,p$ and $k = 1,2,\ldots,n-1$,

$$\alpha_j = n+\lambda_j \quad , \qquad \alpha_{j+kp} = \lambda_j \quad , \tag{2.2.1}$$

and the matrix of associated orthogonal vectors is

$$\underline{Q} = \begin{bmatrix} \frac{\underline{R}}{\sqrt{n}} & \frac{\underline{R}}{\sqrt{2}} & \frac{\underline{R}}{\sqrt{6}} & \cdots & \frac{\underline{R}}{\sqrt{n(n-1)}} \\ \frac{\underline{R}}{\sqrt{n}} & -\frac{\underline{R}}{\sqrt{2}} & \frac{\underline{R}}{\sqrt{6}} & \cdots & \frac{\underline{R}}{\sqrt{n(n-1)}} \\ \frac{\underline{R}}{\sqrt{n}} & \underline{0} & -\frac{2\underline{R}}{\sqrt{6}} & \cdots & \frac{\underline{R}}{\sqrt{n(n-1)}} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \frac{\underline{R}}{\sqrt{n}} & \underline{0} & \underline{0} & \cdots & -\frac{(n-1)\underline{R}}{\sqrt{n(n-1)}} \end{bmatrix} . \tag{2.2.2}$$

Note: Because of the multiplicity of the roots $\{\lambda_j\}$, $\underline{Q}$ is not uniquely defined, but (2.2.2) is a convenient choice. We also note that the form of $\underline{Q}$ will not be used until the next chapter, but it is convenient to give it here.

Proof: To find the characteristic roots of $\underline{A}$, we can solve the determinantal equation $|\underline{A} - \alpha\underline{I}_{np}| = 0$. But $\underline{A} - \alpha\underline{I}_{np} = \underline{I}_n \otimes (\underline{C} - \alpha\underline{I}_p) + \underline{J}_n \otimes \underline{I}_p$. Thus by Lemma 2.2.2, with $\underline{A}$ and $\underline{B}$ of Lemma 2.2.2 being replaced respectively by $\underline{C} - \alpha\underline{I}_p$ and $\underline{I}_p$,

$$|\underline{A} - \alpha\underline{I}_{np}| = |\underline{C} - \alpha\underline{I}_p|^{n-1}|\underline{C} - (\alpha-n)\underline{I}_p| .$$

Thus the solutions of $|\underline{A} - \alpha\underline{I}_{np}| = 0$ are those of $|\underline{C} - \alpha\underline{I}_p| = 0$, each occurring (n-1) times and those of $|\underline{C} - (\alpha-n)\underline{I}_p| = 0$, each once. Since $\underline{R}'\underline{C}\,\underline{R} = \underline{\Lambda}$, where $\underline{\Lambda} = \text{diag}(\lambda_1,\ldots,\lambda_p)$, it follows that the solutions of $|C - (\alpha-n)\underline{I}_p| = 0$ are $n+\lambda_1,\ldots,n+\lambda_p$, and (2.2.1) is proved.

Let $\underline{R} = (\underline{r}_1,\underline{r}_2,\ldots,\underline{r}_p)$. By the definition of characteristic roots and vectors, for $i = 1,2,\ldots,p$, $\underline{C}\,\underline{r}_i = \lambda_i\underline{r}_i$. For $i = 1,2,\ldots,p$ and $k = 1,2,\ldots,n-1$, let q_i and q_{kp+i} be np-variate vectors such that

$$\underline{q}_i' = \frac{1}{\sqrt{n}} (\underline{r}_i', \underline{r}_i', \ldots, \underline{r}_i')$$

$$\underline{q}_{kp+i}' = \frac{1}{\sqrt{k(k+1)}} (\underline{r}_i', \underline{r}_i', \ldots, \underline{r}_i', -k\underline{r}_i', \underline{0}, \ldots, \underline{0}) ,$$

where $-k\underline{r}_i'$ is the $(k+1)$-st p-variate component of $\underline{q}_{kp+i}'$. We must show $\underline{A}\,\underline{q}_i = (n+\lambda_i)\underline{q}_i$ and $\underline{A}\,\underline{q}_{kp+i} = \lambda_i \underline{q}_{kp+i}$. But

$$\underline{A}\,\underline{q}_i = \begin{bmatrix} \underline{I}_p+\underline{C} & \underline{I}_p & \cdots & \underline{I}_p \\ \underline{I}_p & \underline{I}_p+\underline{C} & \cdots & \underline{I}_p \\ \vdots & \vdots & & \vdots \\ \underline{I}_p & \underline{I}_p & \cdots & \underline{I}_p+\underline{C} \end{bmatrix} \begin{bmatrix} \underline{r}_i/\sqrt{n} \\ \underline{r}_i/\sqrt{n} \\ \vdots \\ \underline{r}_i/\sqrt{n} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{n}\underline{r}_i + \underline{C}\,\underline{r}_i/\sqrt{n} \\ \sqrt{n}\underline{r}_i + \underline{C}\,\underline{r}_i/\sqrt{n} \\ \vdots \\ \sqrt{n}\underline{r}_i + \underline{C}\,\underline{r}_i/\sqrt{n} \end{bmatrix} = (n+\lambda_i) \begin{bmatrix} \underline{r}_i/\sqrt{n} \\ \underline{r}_i/\sqrt{n} \\ \vdots \\ \underline{r}_i/\sqrt{n} \end{bmatrix}$$

and

$$\underline{A}\,\underline{q}_{p+i} = \begin{bmatrix} \underline{I}_p+\underline{C} & \underline{I}_p & \cdots & \underline{I}_p \\ \underline{I}_p & \underline{I}_p+\underline{C} & \cdots & \underline{I}_p \\ \vdots & \vdots & & \vdots \\ \underline{I}_p & \underline{I}_p & \cdots & \underline{I}_p+\underline{C} \end{bmatrix} \begin{bmatrix} \underline{r}_i/\sqrt{2} \\ -\underline{r}_i/\sqrt{2} \\ \vdots \\ \underline{0} \end{bmatrix}$$

$$= \begin{bmatrix} \underline{C}\,\underline{r}_i/\sqrt{2} \\ -\underline{C}\,\underline{r}_i/\sqrt{2} \\ \vdots \\ \underline{0} \end{bmatrix} = \lambda_i \begin{bmatrix} \underline{r}_i/\sqrt{2} \\ -\underline{r}_i/\sqrt{2} \\ \vdots \\ \underline{0} \end{bmatrix} .$$

Similarly, for $k = 2,\dots,n-1$, $\underline{A}\,\underline{q}_{kp+i} = \lambda_i \underline{q}_{kp+i}$. Clearly the n-1 vectors corresponding to the root λ_i (having multiplicity n-1) are linearly independent and hence span the subspace corresponding to λ_i. (Of course the chosen $\underline{q}$'s are not the only such set of vectors.) Since $\underline{R}$ is an orthogonal matrix, it is straightforward to show that $\underline{Q}$ is also, and (2.2.2) is proved. □

Theorem 2.2.1. Let $S_{01}, S_{02}, \dots, S_{0n}$ be as defined in paragraph (A) on pages 8 and 9 with $\underline{\mu}_0 = \underline{\mu}$. Then $\overline{S}_0 = n^{-1}\sum_{j=1}^n S_{0j}$ can be represented as

$$\overline{S}_0 = \sum_{j=1}^{p} \{(1+n^{-1}\lambda_j)W_{1j} + n^{-1}\lambda_j W_{2j}\} , \qquad (2.2.3)$$

where $\{W_{ij}, W_{2j}\}_{j=1,2,\dots,p}$ are mutually independent chi-squared random variables, W_{1j} having one degree of freedom and W_{2j} having n-1 degrees of freedom, and $\lambda_1,\dots,\lambda_p$ are the characteristic roots of $\underline{I}_p + 2\underline{\Sigma}^{-1}\underline{\Delta}$.

Proof: By the assumptions on $S_{01},\dots,S_{0n}$, we can express $\overline{S}_0$ as

$$\overline{S}_0 = n^{-1} \sum_{j=1}^{n} \underline{U}'_{0j}\underline{\Sigma}^{-1}\underline{U}_{0j} \quad ,$$

where $\underline{U}_{0j} = \underline{X}_0 - \underline{X}_j + \underline{Y}_{0j} - \underline{Y}_{j0}$ and is distributed $N_p(\underline{0}, 2(\underline{\Sigma}+\underline{\Delta}))$. Let $\underline{\Sigma}^{-\frac{1}{2}}$ be the symmetric square root of $\underline{\Sigma}^{-1}$ (i.e. $\underline{\Sigma}^{-\frac{1}{2}}$ is symmetric and $\underline{\Sigma}^{-\frac{1}{2}}\underline{\Sigma}^{-\frac{1}{2}} = \underline{\Sigma}^{-1}$), and let

$$\underline{U}' = (\underline{U}'_{01}\underline{\Sigma}^{-\frac{1}{2}}, \underline{U}'_{02}\underline{\Sigma}^{-\frac{1}{2}}, \dots, \underline{U}'_{0n}\underline{\Sigma}^{-\frac{1}{2}}) .$$

Thus $\overline{S}_0 = n^{-1}\underline{U}'\underline{U}$ and $\underline{U}$ is distributed normally with mean $\underline{0}$ and covariance matrix

$$\underline{V} = \begin{bmatrix} \underline{V}_{11} & \cdots & \underline{V}_{1n} \\ \vdots & & \vdots \\ \underline{V}_{n1} & \cdots & \underline{V}_{nn} \end{bmatrix}$$

where $\underline{V}_{ij} = E\{\underline{\Sigma}^{-\frac{1}{2}}\underline{U}_{0i}\underline{U}'_{0j}\ \underline{\Sigma}^{-\frac{1}{2}}\}$. Thus $\underline{V}_{ii}$ is the covariance matrix of $\underline{\Sigma}^{-\frac{1}{2}}\underline{U}_{0i}$ or

$$V_{ii} = 2\underline{\Sigma}^{-\frac{1}{2}}(\underline{\Sigma}+\underline{\Delta})\underline{\Sigma}^{-\frac{1}{2}} = 2(\underline{I}_p+\underline{\Omega})$$

where $\underline{\Omega} = \underline{\Sigma}^{-\frac{1}{2}}\underline{\Delta}\ \underline{\Sigma}^{-\frac{1}{2}}$. For $i \neq j$, since $E\underline{X}_0 = E\underline{X}_i = \underline{\mu}$ and $\underline{X}_0, \underline{X}_1, \ldots, \underline{X}_n$ are independent and $\text{cov}(\underline{X}_0) = \underline{\Sigma}$, it follows that

$$V_{ij} = \underline{\Sigma}^{-\frac{1}{2}}E\{\underline{X}_0\underline{X}_0' - \underline{\mu}\ \underline{\mu}'\}\underline{\Sigma}^{-\frac{1}{2}} = \underline{I}_p .$$

Thus

$$\underline{V} = \underline{I}_n \otimes (\underline{I}_p + 2\Omega) + \underline{J}_n \otimes \underline{I}_p .$$

But if $\overline{S}_0 = n^{-1}\underline{U}'\underline{U}$ where $\underline{U}$ is $N_{np}(\underline{O},\underline{V})$, then we can represent $\overline{S}_0$ as

$$\overline{S}_0 = n^{-1}\sum_{j=1}^{np} \alpha_j Z_j^2 \qquad (2.2.4)$$

where $Z_1,\ldots,Z_{np}$ are independent unit normal random variables and $\alpha_1,\ldots,\alpha_{np}$ are the characteristic roots of $\underline{V}$ (see, e.g., Johnson and Kotz [11], chapter 29, section 5). By Lemma 2.2.3, $\alpha_1,\ldots,\alpha_{np}$ consist of $n+\lambda_j$, each occurring once, and λ_j, each occurring n-1 times, where $\lambda_1,\ldots,\lambda_p$ are the characteristic roots of $\underline{I}_p + 2\underline{\Omega}$, or equivalently of $\underline{I}_p + 2\underline{\Sigma}^{-1}\underline{\Delta}$. The result follows immediately on substitution in (2.2.4). □

Since $\overline{S}_0$ can be represented in the form given in (2.2.4), it is clearly a quadratic form in normal variables. Consequently, we would expect that evaluation of quantities like $\Pr(\overline{S}_0 > s)$ would involve straightforward application of one of the well-known methods such as the expansions described in chapter 29 of Johnson and Kotz [11]. Unfortunately, that is not the case. The reason for this is that, for reasonably rapid convergence of these expansions, the coefficients (the

α's in (2.2.4)) should be about the same size. In the situation we have, it would be reasonable to assume that the measurement error made in determining the distances $\underline{S}_{0j}$ is small relative to the overall variability in the problem; otherwise, there would be little hope of drawing meaningful conclusions. Thus the characteristic roots of $\underline{\Sigma}^{-1}\underline{\Delta}$ can be assumed to be small, certainly no larger than one, which in turn implies that $\lambda_1,\ldots,\lambda_p$ are about one or a little larger. Thus p of the coefficients in (2.2.4) are close to one and the other (n-1)p are close to zero. Because of this, the expansions mentioned above converge too slowly to be of any practical use.

However, $\Pr(\overline{S}_0 > s)$ gives us the probability of misclassifying an individual which is truly from the population, a quantity which we wish to be able to compute. Before examining the problem of evaluation of $\Pr(\overline{S}_0 > s)$ in general, let us consider a special case. Specifically, suppose that $\underline{\Sigma}$ and $\underline{\Delta}$ are related by $\underline{\Delta} = \delta\underline{\Sigma}$, where $0 \le \delta < 1$; that is the variability of the measurement error is similar to the population variability, but smaller in magnitude. Since, as mentioned above, we may assume that the measurement error is small compared to the population variability, restricting $\delta < 1$ would seem to be reasonable. The additional restriction that $\underline{\Delta}$ be a constant multiple of $\underline{\Sigma}$ may be valid in some cases, and it leads to considerable simplification of the distribution in question. Under these conditions, $\underline{\Sigma}^{-1}\underline{\Delta} = \delta\underline{I}_p$ and we have $\lambda_1=\ldots=\lambda_p = 1 + 2\delta$. Let us denote the common value by λ. The representation (2.2.3) then becomes, by reproductivity of chi-squared random variables,

$$\overline{S}_0 = (1 + n^{-1}\lambda)W_1 + n^{-1}\lambda W_2 , \tag{2.2.5}$$

where W_1 and W_2 are independent chi-squared random variables with p and (n-1)p degrees of freedom respectively. We now wish to obtain an expression for $\Pr(\overline{S}_0 > s)$ when $\overline{S}_0$ has the form (2.2.5). First we prove a lemma.

Lemma 2.2.4. Let $p \geq 2$ be an even integer, ν an arbitrary positive integer and a,b be arbitrary positive constants such that $a > b$; let X and Y be distributed independently as chi-squared random variables with p and ν degrees of freedom respectively, and let $Z = aX + bY$. Then

$$\Pr(Z > z) = 1 - F_\nu(\tfrac{z}{b}) + (\tfrac{a}{a-b})^{\frac{1}{2}\nu} \exp\{-\tfrac{z}{2a}\}$$
$$\times \left\{ \sum_{\ell=0}^{q} (-\tfrac{b}{a-b})^{\ell} (\tfrac{1}{2}\nu)_\ell h_\ell^{(q)}(\tfrac{z}{2a}) F_{\nu+2\ell}(\tfrac{a-b}{ab} z) \right\}, \tag{2.2.6}$$

where $F_\nu(x)$ is the cumulative distribution function of an χ^2_ν random variable, $q = \frac{1}{2}p-1$, $(x)_a = \Gamma(x+a)/\Gamma(x)$, and for $\ell = 0,1,\ldots,q$,

$$h_\ell^{(q)}(x) = \frac{1}{\ell!} \sum_{k=0}^{q-\ell} \frac{x^k}{k!} . \tag{2.2.7}$$

__Proof__: By definition, we have

$$\Pr(Z > z) = \Pr(aX+bY > z) = \Pr(Y > \tfrac{z}{b}) + \Pr(Y \leq \tfrac{z}{b}, X > \tfrac{z-bY}{a})$$
$$= 1 - F_\nu(\tfrac{z}{b}) + \int_{t=0}^{b^{-1}z} \Pr(X > \tfrac{z-bt}{a}) f_\nu(t)dt , \tag{2.2.8}$$

where

$$f_\nu(t) = (\tfrac{1}{2}t)^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}t}/\Gamma(\tfrac{1}{2}\nu)$$

is the probability density function of a x^2_ν random variable. But for p even and $q = \frac{1}{2}p-1$,

$$\Pr(X > x) = e^{-x/2} \sum_{j=0}^{q} (\tfrac{1}{2}x)^j / j! \ .$$

Hence, letting $I(z)$ be the integral in (2.2.8), we have

$$I(z) = \int_{t=0}^{b^{-1}z} \left\{\exp\left(-\frac{z-bt}{a}\right)\right\}\left\{\sum_{j=0}^{q} \left(\frac{z-bt}{a}\right)^j / j!\right\} f_\nu(t)\,dt$$

$$= e^{-z/2a} \sum_{j=0}^{q} \int_{t=0}^{b^{-1}z} (j!)^{-1}\left\{\sum_{\ell=0}^{j} \binom{j}{\ell}\left(\frac{z}{2a}\right)^{j-\ell}\left(-\frac{bt}{2a}\right)^{\ell}\right\}$$

$$\times \left\{\tfrac{1}{2}(\tfrac{1}{2}t)^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}t} / \Gamma(\tfrac{1}{2}\nu)\right\} dt$$

$$= e^{-z/2a} \sum_{j=0}^{q} \sum_{\ell=0}^{j} [\ell!(j-\ell)!]^{-1} \left(\frac{z}{2a}\right)^{j-\ell} \left(-\frac{b}{a}\right)^{\ell}$$

$$\times \int_{t=0}^{b^{-1}z} (\tfrac{1}{2})(\tfrac{1}{2}t)^{\frac{1}{2}\nu+\ell-1} \exp\{-\tfrac{1}{2}t(1-ba^{-1})\}/\Gamma(\tfrac{1}{2}\nu)\,dt \ .$$

Making the change of variable $u = t(1 - ba^{-1})$ and rearranging terms, we obtain

$$I(z) = e^{-z/2a} \sum_{\ell=0}^{q} (-ba^{-1})^{\ell} \left\{\sum_{j=\ell}^{q} [\ell!(j-\ell)!]^{-1} (z/2a)^{j-\ell}\right\}$$

$$\times \left(\frac{a}{a-b}\right)^{\frac{1}{2}\nu+\ell} \int_{u=0}^{z((a-b)/(ab))} (\tfrac{1}{2})(\tfrac{1}{2}u)^{\frac{1}{2}\nu+\ell-1} e^{-\frac{1}{2}u} / \Gamma(\tfrac{1}{2}\nu)\,du \ . \qquad (2.2.9)$$

But the integral in (2.2.9) is $(\tfrac{1}{2}\nu)_\ell F_{\nu+2\ell}(\frac{a-b}{ab} z)$, and the term in braces is $h_\ell^{(q)}(x)$ as defined in (2.2.7) (with an index shift). Hence the result follows immediately on substitution and cancellation of terms in (2.2.9). $\square$

Theorem 2.2.2. If p is even, and $S_{01},\ldots,S_{0n}$ are as defined in paragraph (A) on pages 8-9, with $\underline{\mu}_0 = \underline{\mu}$, $\underline{\Delta} = \delta\underline{\Sigma}$, and $\bar{S}_0 = n^{-1}\sum_{j=1}^{n} S_{0j}$, then

$$P(\bar{S}_0>s) = 1 - F_\nu(ns\lambda^{-1}) + \exp\{-\tfrac{1}{2}s/(1+n^{-1}\lambda)\}(1+n^{-1}\lambda)^{\frac{1}{2}\nu}$$
$$\times \{\sum_{\ell=0}^{q} (-n^{-1}\lambda)^\ell (\tfrac{1}{2}\nu)_\ell h_\ell^{(q)}(\tfrac{1}{2}s/(1+n^{-1}\lambda))F_{\nu+2\ell}(sn^2/n\lambda+\lambda^2))\},$$
$$(2.2.10)$$

where $q = \frac{1}{2}p-1$, $\nu = p(n-1)$, $\lambda = 1+2\delta$, $F_\nu(x)$ is the cumulative distribution function of a χ_ν^2 random variable, and $h_\ell^{(q)}(x)$ is a polynomial of degree q in x, as defined in (2.2.7).

Proof: We have already shown that, under the conditions of this theorem, $\bar{S}_0$ can be represented as in (2.2.5), that is

$$\bar{S}_0 = (1+n^{-1}\lambda)W_1 + n^{-1}\lambda W_2 ,$$

where W_1 and W_2 are independent chi-squared random variables with p and p(n-1) degrees of freedom respectively and $\lambda = 1+2\delta$. Thus the proof follows directly from Lemma 2.2.4 by substitution of $b = n^{-1}\lambda$, $a = 1+n^{-1}\lambda$ and $\nu = p(n-1)$. Here a-b = 1, allowing simplification of the expression (2.2.6) to the form given here. □

While the expression (2.2.10) appears to be rather complicated, the values of $F_{\nu+2\ell}(\cdot)$ are tabulated or can be easily computed or approximated, and, for dimensions of practical interest, the sum in (2.2.10) has very few terms. Thus, even though the peculiar form of the linear combination of chi-squared random variables with which we are confronted makes the usual computational methods of little or no use, that form can also be exploited, as in Theorem 2.2.2, to give a practical means of computing the probabilities of interest, at least when p is even.

For odd values of p, there does not seem to be a simple closed form for the probability. It is true that if X is a chi-squared random variable with p = 2q-1 degrees of freedom (q even), then

$$P(X>x) = e^{-\frac{1}{2}x}\left\{\sum_{j=0}^{q-1} \frac{(\frac{1}{2}s)^{j+\frac{1}{2}}}{\Gamma(j+3/2)}\right\} + 2[1 - \Phi(\sqrt{x})],$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, and this expression does allow us to write an expression expanding (2.2.8) in this case. However, here the term in braces above, instead of resulting in a linear combination of chi-squared probabilities, results in a more general combination of confluent hypergeometric functions. In addition, an integral involving $\Phi(\cdot)$ results, and it does not appear to have a closed form. It could, of course, be evaluated numerically.

Because of the representation of $\overline{S}_0$ as

$$\overline{S}_0 = \sum_{j=1}^{p} (1 + n^{-1}\lambda_j)W_{1j} + \sum_{j=1}^{p} n^{-1}\lambda_j W_{2j} ,$$

with W_{1j} and W_{2j} the appropriate chi-squared random variables, and the λ's fixed constants, if n is large, the first of the components above is distributed much like a χ_p^2 random variable and the second component behaves much like a constant. In fact it is simple to show that to be the limiting distribution as $n \to \infty$.

Theorem 2.2.3. If $\lambda_1,\ldots,\lambda_p$ and $\overline{S}_0$ are as given in Theorem 2.2.1, then the limiting distribution of $\overline{S}_0$ as $n \to \infty$ is $U + \sum_{j=1}^{p}\lambda_j$, where U is distributed as χ_p^2.

Proof: The characteristic function of a χ^2_ν random variable is $(1 - 2it)^{-\frac{1}{2}\nu}$. Thus the characteristic function of $\bar{S}_0$ is

$$\phi(t) = \prod_{j=1}^{p} \{[1 - 2it(1+n^{-1}\lambda_j)]^{-\frac{1}{2}}[1 - 2itn^{-1}\lambda_j]^{-\frac{1}{2}(n-1)}\}$$

$$= \prod_{j=1}^{p} A_j(t)B_j(t)C_j(t) \qquad (2.2.11)$$

where $A_j(t) = (1-2it-2itn^{-1}\lambda_j)^{-\frac{1}{2}}$, $B_j(t) = (1-2itn^{-1}\lambda_j)^{-\frac{1}{2}n}$, and $C_j(t) = (1-2itn^{-1}\lambda_j)^{\frac{1}{2}}$. But as $n \to \infty$, $A_j(t) \to (1-2it)^{-\frac{1}{2}}$, $B_j(t) \to \exp(it\lambda_j)$ and $C_j(t) \to 1$. Thus as $n \to \infty$,

$$\phi(t) \to (1-2it)^{-\frac{1}{2}p}\exp(it\textstyle\sum_{j=1}^{p}\lambda_j) ,$$

which is the characteristic function of $U + \sum_{j=1}^{p}\lambda_j$, and the result follows by the uniqueness of characteristic functions. □

Of course, the limiting distribution gives us an approximation to the distribution of $\bar{S}_0$ for large n. Another condition which leads to a relatively simple approximation is if $\underline{\Delta}$ is small. We saw earlier that if $\underline{\Delta} = \underline{0}$, that is, when there is no distortion from measurement (or perception) error, the test based on $\bar{S}_0$ is approximately equivalent to the two sample likelihood ratio test for equality of means. In practical situations, $\underline{\Delta}$ may be small compared to Σ, and that limiting distribution as $\underline{\Delta} \to \underline{0}$ can be used as an approximation. We already have the characteristic function of $\bar{S}_0$ in (2.2.11). But as $\underline{\Delta} \to \underline{0}$, $\lambda_j \to 1$ for all j, and hence as $\underline{\Delta} \to \underline{0}$,

$$\phi(t) \to [1-2it(1+n^{-1})]^{-\frac{1}{2}p}[1-2itn^{-1}]^{-\frac{1}{2}p(n-1)} .$$

By inversion of the characteristic function, the distribution of $\bar{S}_0$ tends, as $\underline{\Delta} \to \underline{0}$, to that of $(1+n^{-1})U + n^{-1}V$, where U and V are independent

and distributed as χ^2_p and $\chi^2_{(n-1)p}$ respectively. This distribution has already been given: it is the special case in Theorem 2.2.2 with $\lambda_1 = \ldots = \lambda_p = 1$.

2.3 Asymptotic expansion for distribution of $\overline{S}_0$.

While the exact distribution of $\overline{S}_0$ is difficult to compute in general, the limiting distribution as $n \to \infty$ is quite simple and is easily computed. Thus it is of interest to develop an asymptotic expansion for the distribution, both to give an indication of the rate of convergence to the limit distribution and to provide an approximation to the distribution which is easier to compute than the exact distribution. In this section we shall develop such an expansion. First, however, we need two lemmas.

Lemma 2.3.1. Let X be a chi-squared random variable with p degrees of freedom, and let Y be a non-negative random variable (depending on n), independent of X, with $EY = O(1)$ as $n \to \infty$ and with r-th central moment, μ_r, of order $O(n^{-r+[r/2]})$ for all integers $r \geq 2$, $[t]$ denoting the integer part of t. Further, let $S = X+Y$ and μ^* denote a bound for EY. Then for $s > \mu^*$, $r \geq 1$,

$$\Pr(S>s) = \Pr(X>s - EY) + \sum_{j=2}^{2r} \frac{c_j \mu_j}{j!} + o(n^{-r}) , \tag{2.3.1}$$

where $c_j = \frac{d^{j-1}}{dy^{j-1}} f_p(s-y)\Big|_{y=EY}$ and $f_p(x)$ is the density function of X.

With c_0 arbitrary, the coefficients $\{c_j\}_{j=0,1,2,\ldots}$ satisfy the following recurrence relation:

$$c_{j+2} = -\tfrac{1}{2} \left\{ \frac{jc_j}{s-EY} + \left(\frac{p-2j-2}{s-EY} - 1\right) c_{j+1} \right\} . \tag{2.3.2}$$

Proof: Throughout the proof, Y and its distribution and moments depend on n, but that dependence is suppressed notationally for simplicity. Let G(y) be the distribution function of Y, $\mu = EY$, and

$$h_s(y) = \Pr(X>s-y) = \int_{s-y}^{\infty} f_p(x)dx .$$

Then

$$\Pr(S>s) = E[h_s(Y)] = \int_0^{\infty} h_s(y)dG(y) .$$

We wish to expand $h_s(y)$ about $y = \mu$, using Taylor's formula and integrate the result. However, $h_s(y)$ may not satisfy the necessary conditions on continuity and differentiability over the entire range of integration. Specifically, the j-th derivative with respect to y, $h_s^{(j)}(y)$, has a discontinuity at $y = s$ if $j \geq \frac{1}{2}(p-2)$. To avoid these difficulties, let b be a fixed constant such that $\mu^* < b < s$. Since $b - \mu^* > 0$ (independent of n), we have by a Chebyshev-type inequality,

$$\Pr(Y>b) \leq \Pr(|Y-\mu| > b-\mu^*) \leq \frac{\mu_{2r+2}}{(b-\mu^*)^{2r+2}} .$$

But for $r \geq 1$, $\mu_{2r+2} = o(n^{-r})$. Thus $\Pr(Y>b) = o(n^{-r})$ for any $r \geq 1$.

On the interval $0 \leq y \leq b$, $h_s(y)$ is everywhere differentiable an arbitrary number of times for any $p \geq 1$. Thus on that interval, we can expand $h_s(y)$ using Taylor's formula, obtaining for $r \geq 1$,

$$h_s(y) = h_s(\mu) + \sum_{j=1}^{2r} (y-\mu)^j h_s^{(j)}(\mu)/j!$$

$$+ (y-\mu)^{2r+1} h_s^{(2r+1)}(\xi)/(2r+1)! , \qquad (2.3.3)$$

where ξ is between y and μ and therefore also $0 \leq \xi < b$. Since

$$\Pr(S>s) = \int_0^b h_s(y)\,dG(y) + \int_b^\infty h_s(y)\,dG(y) \tag{2.3.4}$$

and

$$\mu_j = \int_0^b (y-\mu)^j\,dG(y) + \int_b^\infty (y-\mu)^j\,dG(y) \ , \tag{2.3.5}$$

substitution of (2.3.3) and (2.3.5) in (2.3.4), together with the fact that $\mu_1 = 0$, yields

$$\Pr(S>s) = h_s(\mu) + \sum_{j=2}^{2r} c_j\mu_j/j! + R_r \ , \tag{2.3.6}$$

where

$$\begin{aligned} R_r &= \left[\int_0^b \frac{(y-\mu)^{2r+1}}{(2r+1)!} h_s^{(2r+1)}(\xi)\,dG(y)\right] \\ &+ \left[\int_b^\infty h_s(y)\,dG(y) - \int_b^\infty h_s(\mu)\,dG(y)\right] \\ &- \left[\sum_{j=1}^{2r} \frac{h_s^{(j)}(\mu)}{j!} \int_b^\infty (y-\mu)^j\,dG(y)\right] \\ &= T_1 + T_2 - T_3 \text{ (say)}, \end{aligned}$$

and $c_2,\ldots,c_{2r}$ are as defined in the theorem. Because $\mu^* < b < s$, $|h_s^{(2r+1)}(y)|$ is bounded for $0 < y < b$, by M say.

Let ν_r be the r-th absolute central moment of Y. By Liapounoff's inequality (see, e.g., Kendall and Stuart [13], p. 62), $\nu_{2r+1}^2 \le \nu_{2r}\nu_{2r+2}$. Thus since, for $r \ge 1$, $\nu_{2r} = O(n^{-r})$ and $\nu_{2r+2} = O(n^{-r-1})$, we have $\nu_{2r+1} = O(n^{-r-\frac{1}{2}})$. Therefore it follows that for $r \ge 1$,

$$|T_1| \le \int_b^\infty |y-\mu|^{2r+1}\,dG(y) \le \nu_{2r+1} = o(n^{-r}) \ .$$

Also, since $0 \le h_s(y) = P(X>s-y) \le 1$ for all y,

$$|T_2| \le 2\int_b^\infty dG(y) = o(n^{-r}) .$$

For T_3, let $T_{3j} = \int_b^\infty (y-\mu)^j dG(y)$ and let

$$u(y) = \begin{cases} 1 \text{ if } |y-\mu| \le 1 \\ 0 \text{ if } |y-\mu| > 1 \end{cases} .$$

Then for $r \ge 1$,

$$\left| \int_b^\infty u(y)(y-\mu)^j \right| dG(y)$$

$$\le \int_b^\infty dG(y) = o(n^{-r})$$

and for $1 \le j \le 2r$, $r \ge 1$,

$$\left| \int_b^\infty [1-u(y)](y-\mu)^j dG(y) \right|$$

$$\le \int_b^\infty |y-\mu|^{2r+1} dG(y) \le \nu_{2r+1}$$

$$= o(n^{-r}) .$$

Hence

$$|T_{3j}| \le \left| \int_b^\infty u(y)(y-\mu)^j dG(y) \right|$$

$$+ \left| \int_b^\infty [1-u(y)](y-\mu)^j dG(y) \right|$$

$$= o(n^{-r}) ,$$

and

$$|T_3| \le \sum_{j=1}^{2r} |h_s^{(j)}(\mu)T_{3j}|/j! = o(n^{-r}) .$$

Thus $|R_r| \le |T_1| + |T_2| + |T_3| = o(n^{-r})$ and (2.3.1) is proved.

To prove the recurrence relation, let $z = \frac{1}{2}(s-y)$. Then

$$h_s^{(r)}(y) = \frac{d^r}{dy^r} f_p(s-y) = \frac{d}{dz}\left[\frac{d^{r-1}}{dy^{r-1}} f_p(s-y)\right]\frac{dz}{dy} .$$

It follows by induction, since $\frac{dz}{dy} = -\frac{1}{2}$, that

$$\frac{d^r}{dy^r} f_p(2z) = (-\tfrac{1}{2})^r \frac{d^r}{dz^r} f_p(2z) .$$

Now $f_p(2z) = \frac{1}{2}z^{\frac{1}{2}p-1}e^{-z}/\Gamma(\frac{1}{2}p)$, and

$$\frac{d^r}{dz^r}[z^{\frac{1}{2}p-1}e^{-z}] = r![L_r^{(\frac{1}{2}p-r-1)}(z)]z^{\frac{1}{2}p-r-1}e^{-z} ,$$

where

$$L_r^{(\gamma)}(z) = \sum_{j=0}^{r} \binom{r+\gamma}{r-j}\frac{(-z)^j}{j!} , \qquad (2.3.7)$$

the generalized Laguerre polynomial, which, while properly defined only for $\gamma > -1$, can be defined formally as in (2.3.7) for all γ, using the convention that $\binom{r}{k} = 0$ if r is a positive integer and $k > r$ or if k is a negative integer.

Thus we have

$$\frac{d^r}{dy^r} f_p(s-y) = (-\tfrac{1}{2})^r r! L_r^{(\frac{1}{2}p-r-1)}(z)z^{-r}f_p(2z) .$$

We can now use known relations, involving Laguerre polynomials (see, e.g., Gradshteyn and Ryzhik [7], p. 1037) to obtain the desired recurrence relation. In particular, we use the three relations

$$(n+1)L_{n+1}^{(\alpha)}(z) = (2n+\alpha+1-z)L_n^{(\alpha)}(z) - (n+\alpha)L_{n-1}^{(\alpha)}(z)$$

$$(n+\alpha)L_{n-1}^{(\alpha)}(z) = nL_n^{(\alpha)}(z) + zL_{n-1}^{(\alpha)}(z)$$

$$L_{n+1}^{(\alpha)}(z) = L_{n+1}^{(\alpha+1)}(z) - L_n^{(\alpha)}(z) .$$

Combining these yields

$$(n+1)L_{n+1}^{(\alpha-1)}(z) = (\alpha-z)L_n^{(\alpha)}(z) - zL_{n-1}^{(\alpha+1)}(z) ,$$

or putting $n = j$ and $\alpha = \frac{1}{2}p - j - 1$,

$$(j+1)L_{j+1}^{(\frac{1}{2}p-j-2)}(z) = (\tfrac{1}{2}p-j-1-z)L_j^{(\frac{1}{2}p-j-1)}(z) - zL_{j-1}^{(\frac{1}{2}p-j)}(z). \qquad (2.3.8)$$

Thus, applying (2.3.8) to the definition of c_{j+2}, we have

$$c_{j+2} = \frac{d^{j+1}}{dy^{j+1}} f_p(s-y)\Big|_{y=EY} ,$$

$$\frac{d^{j+1}}{dy^{j+1}} f_p(s-y) = (-\tfrac{1}{2})^{j+1}(j+1)!L_{j+1}^{(\frac{1}{2}p-j-2)}(z)z^{-(j+1)}f_p(2z)$$

$$= (\tfrac{1}{2})^{j+1}(j+1)!\left[\frac{\frac{1}{2}p-j-1-z}{j+1} L^{(\frac{1}{2}p-j-1)}(z) - zL_{j-1}^{(\frac{1}{2}p-j)}(z)\right]z^{-j-1}f_p(2z)$$

$$= (-\tfrac{1}{2})\left(\frac{\frac{1}{2}p-j-1}{z} -1\right)(-\tfrac{1}{2})^j j!L_j^{(\frac{1}{2}p-j-1)}(z)z^{-j}f_p(2z)$$

$$- (\tfrac{1}{2})^2 \frac{j(j-1)!}{z}(-\tfrac{1}{2})^{j-1}L_{j-1}^{(\frac{1}{2}p-j)}(z)z^{-j+1}f_p(2z)$$

$$= -\tfrac{1}{2}\left[\frac{p-2j-2}{2z} -1\right]\frac{d^j}{dy^j} f_p(2z) -\tfrac{1}{2}\left(\frac{j}{2z}\right)\frac{d^{j-1}}{dy^{j-1}} f_p(2z) .$$

Evaluating this expression at $y = EY$ then yields

$$c_{j+2} = -\tfrac{1}{2}\left[\left(\frac{p-2j-2}{s-EY} -1\right)c_{j+1} + \frac{jc_j}{s-EY}\right] .$$

c_1 is defined to be $f_p(s-EY)$, and we can define c_0 arbitrarily and this relation will still hold for $j = 0$. Thus we have obtained the desired relation (2.3.2). We note that (2.3.2) can also be derived directly, using Leibnitz' rule for the derivative of a product. □

Lemma 2.3.2. Let Y be a random variable (depending on n) with all moments and cumulants existing and such that the r-th cumulant, κ_r, is of order $O(n^{1-r})$ for all $r \geq 2$. Then for $r \geq 2$, the r-th central moment of Y, μ_r, is of order $O(n^{-r+[r/2]})$, [t] denoting the integer part of t.

Proof: It is well-known that, subject to conditions of existence, we can write

$$\mu_r = \sum_{j_1} \sum_{j_2} \cdots \sum_{j_\ell} a_{j_1} a_{j_2} \ldots a_{j_\ell} \kappa_{j_1} \kappa_{j_2} \ldots \kappa_{j_\ell} ,$$

where $j_i \geq 2$, $\sum_{i=1}^{\ell} j_i = r$, and the a_j's are constants not depending on the distribution of Y (see Kendall and Stuart [13], p. 70). But $\kappa_{j_i} = O(n^{1-j_i})$. Hence

$$\mu_r = O\left(n^{\Sigma_{i=1}^{\ell}(1-j_i)}\right) = O(n^{m-r}) ,$$

where m is the maximum number of cumulants in any term of the sum. Clearly there is always a term involving [r/2] cumulants, namely with $j_1 = j_2 = \ldots = j_{r/2} = 2$ if r is even and with $j_1 = 3$, $j_2 = \ldots = j_{[r/2]} = 2$ if r is odd, and no terms involving more. Thus $m = [r/2]$. □

Theorem 2.3.1. Let $S_{01}, S_{02}, \ldots, S_{0n}$ be as defined in paragraph (A) on pages 8-9 with $\mu_0 = \mu$, and $\overline{S}_0 = n^{-1}\sum_{j=1}^{n} S_{0j}$; let $\lambda_1, \ldots, \lambda_p$ be the characteristic roots of $\underline{I}_p + 2\underline{\Sigma}^{-1}\underline{\Delta}$ and

$$V = n^{-1}\sum_{j=1}^{p}\lambda_j W_{2j} ,$$

where $W_{21},\ldots,W_{2p}$ are independent chi-squared random variables with n-1 degrees of freedom. Then for $s > \sum_{j=1}^{p}\lambda_j$, $r \geq 1$,

$$\Pr(S_0 > s) = d_0\Pr[\chi_p^2 > (s-EV)/\beta] + \sum_{k=1}^{r}\Bigg\{d_k\Pr[\chi_{p+2k}^2 > (s-EV)/\beta] + \sum_{j=1}^{k} d_{k-j}\left[\frac{a_{k-j,2j-1}\mu_{2j-1}}{(2j-1)!} + \frac{a_{k-j,2j}\mu_{2j}}{(2j)!}\right]\Bigg\} + o(n^{-r}) , \tag{2.3.9}$$

where

$$\beta = 1+n^{-1}\bar{\lambda} , \quad \bar{\lambda} = p^{-1}\sum_{j=1}^{p}\lambda_j , \quad \mu_j = E[(V-EV)/\beta]^j ,$$

$$d_k = (-1)^k\sum_{j=k}^{r}\binom{j}{k}c_j\beta^{-j} \qquad (k = 0,1,\ldots,r)$$

$$c_0 = 1$$

$$c_k = (2k)^{-1}\sum_{j=0}^{k-1}B_{k-j}c_j \qquad (k = 1,2,\ldots,r)$$

$$B_k = \sum_{j=1}^{p}\left(\frac{\bar{\lambda}-\lambda_j}{n}\right)^k \qquad (k = 0,1,\ldots,r)$$

and $\{a_{k,j}\}_{k=0,1,\ldots;j=0,1,\ldots}$ satisfy the following relations:

$$a_{k,0} = 1 \qquad (k = 0,1,\ldots)$$

$$a_{0,1} = \tfrac{1}{2}(\tfrac{1}{2}(s-EV)/\beta)^{\frac{1}{2}p-1}\exp\{-\tfrac{1}{2}(s-EV)/\beta\}/\Gamma(\tfrac{1}{2}p)$$

$$a_{k+1,1} = a_{k,1}\left\{\frac{s-EV}{(p+2k)\beta}\right\} \qquad (k = 0,1,\ldots)$$

and for any $k \geq 0$,

$$a_{k,j+2} = -\tfrac{1}{2}\left\{\left[\frac{(p+2k-2j-2)\beta}{s-EV} - 1\right]a_{k,j+1} + \left[\frac{j\beta}{s-EV}\right] a_{k,j}\right\} \quad (j = 0,1,\ldots) .$$

<u>Proof</u>: By Theorem 2.2.1, we have that

$$\overline{S}_0 = \sum_{j=1}^{p} (1+n^{-1}\lambda_j)W_{1j} + \sum_{j=1}^{p} n^{-1}\lambda_j W_{2j} = U + V$$

where W_{1j} is χ_1^2, W_{2j} is χ_{n-1}^2 and the W's are independent. Thus U is a quadratic form in normal variables and for n sufficiently large, we have $n > \max_{1\le j\le p} (\lambda_j - 2\overline{\lambda})$. This implies $1 + n^{-1}\overline{\lambda} > 2(1+n^{-1}\lambda_{max})$, and we can express the distribution function G(u) of U using a Laguerre series expansion as described in Johnson and Kotz [11], chapter 29, section 5:

$$1 - G(u) = \sum_{j=0}^{\infty} c_j \beta^{-j} j! [\Gamma(\tfrac{1}{2}p)/\Gamma(\tfrac{1}{2}p+j)]$$

$$\times \int_u^{\infty} f_p(\beta^{-1}x) L_j^{(\frac{1}{2}p-1)}(\tfrac{1}{2}\beta^{-1}x)dx , \qquad (2.3.10)$$

where $\beta = 1+n^{-1}\overline{\lambda}$, $f_p(x)$ is the density function of a χ_p^2 random variable and

$$c_0 = 1$$

$$c_k = (2k)^{-1} \sum_{j=0}^{k-1} B_{k-j} c_j \qquad (k \ge 1)$$

$$B_k = \sum_{j=1}^{p} \left(\frac{\overline{\lambda}-\lambda_j}{n}\right)^k \qquad (k \ge 0)$$

and $L_j^{(\alpha)}(x)$ is the Laguerre polynomial defined in (2.3.7). By substitution of the definition of $L_j^{(\alpha)}(x)$, we can write (2.3.10) in the equivalent form

$$1 - G(u) = \sum_{j=0}^{r} c_j \beta^{-j} \sum_{k=0}^{j} (-1)^k \binom{j}{k} Pr(\chi_{p+2k}^2 > \beta^{-1}u) + R_r(u) , \qquad (2.3.11)$$

where $R_r(u)$ is the sum of the remaining terms in the infinite sum. Gurland [9] has shown that

$$|R_r(u)| \le M \max_{1\le j\le p} \left|\frac{\bar{\lambda}-\lambda_j}{n\beta}\right|^{r+1} = O(n^{-r-1}) .$$

Alternatively, we may observe that for $\max_{1\le j\le p} |\bar{\lambda}-\lambda_j| < 1$, $|B_k| \le pn^{-k}$. Hence for $k \ge 1$,

$$|c_k| < \frac{k+1}{4k}\left(\frac{p}{n}\right)^k \le \tfrac{1}{2}\left(\frac{p}{n}\right)^k .$$

Thus for sufficiently large n,

$$|R_r(u)| \le \sum_{j=r+1}^{\infty} |c_j|\,|\beta^{-j}| \sum_{k=0}^{j} \binom{j}{k} \Pr(\chi^2_{p+2k} > \beta^{-1}u)$$

$$\le \tfrac{1}{2}\left(\frac{p}{n}\right)^{r+1} \sum_{j=0}^{\infty} \left(\frac{2p}{n\beta}\right)^j = o(n^{-r}) .$$

Upon switching the order of summation in (2.3.11), we obtain

$$1 - G(u) = \sum_{k=0}^{r} dk \Pr(\chi^2_{p+2k} > \beta^{-1}u) + o(n^{-r}),$$

where $d_k = (-1)^k \sum_{j=k}^{r} \binom{j}{k} c_j$. Therefore

$$\Pr(\bar{S}_0 > s) = \Pr(U+V>s)$$

$$= \sum_{k=0}^{r} d_k \Pr(\chi^2_{p+2k} + \beta^{-1}V > \beta^{-1}s) + o(n^{-r}) . \qquad (2.3.12)$$

But $V = n^{-1}\sum_{j=1}^{p} \lambda_j W_{2j}$, so the m-th cumulant of V is

$$\kappa_m(V) = n^{-m} \sum_{j=1}^{p} \lambda_j^m k_m(W_{2j})$$

and hence, since W_{2j} is a χ^2_{n-1} random variable,

$$\kappa_m(\beta^{-1}V) = (\sum_{j=1}^{p} \lambda_j^m)2^{m-1}(m-1)!(n-1)n^{-m}\beta^{-m} = O(n^{1-m}) .$$

Thus V/β satisfies the conditions of Lemma 2.3.2 and the m-th central moment of V/β, denoted by μ_m is of order $O(n^{-m+[m/2]})$, satisfying the conditions of Lemma 2.3.1. Therefore, for each k, we have for $s > \sum_{j=1}^{p}\lambda_j > EV$,

$$\Pr(\chi^2_{p+2k} + \beta^{-1}V > \beta^{-1}s) = \Pr(\chi^2_{p+2k} > (s-EV)/\beta)$$

$$+ \sum_{j=2}^{2r} a_{k,j}\mu_j/j! + o(n^{-r}) , \qquad (2.3.13)$$

where $a_{k,0} = 1$, $a_{k,1} = f_{p+2k}((s-EV)/\beta)$ and $\{a_{k,j}\}_{k=0,1,\ldots,j=0,1,\ldots}$ satisfy the stated recurrence relation. Also it is easily verified that

$$a_{k+1,1} = \frac{(s-EV)}{\beta(p+2k)} a_{k,1}$$

as required. The exact form of (2.3.9) results from substitution of (2.3.13) in (2.3.12), omitting terms of order $o(n^{-r})$, determined by the fact that $d_k = o(n^{-k+1})$, $\mu_2 = O(n^{-1})$ and for $j \geq 2$, μ_{2j-1} and μ_{2j} are both $O(n^{-j})$. □

Since (2.3.9) is quite a complicated expression, it would be hoped that a good approximation to the distribution of $\overline{S}_0$ be obtained by including only a few terms. In the next section we shall examine the accuracy of (2.3.9) as an approximation and find that often this is the case. Several remarks should be made here concerning portions of (2.3.9) which are actually not quite as complicated as they appear there:

<u>Remark 1</u>. With the particular choice of $\beta = 1+n^{-1}\overline{\lambda}$, we have $B_1 = 0$ and $c_0 = 0$, simplifying some of the formulas for the coefficients.

Remark 2. Since $\mu_1 = 0$, the first term in the inner sum (j=1) is really

$$d_{k-1}a_{k-1,2}\mu_2/2!$$

but it was left in the more general form for notational convenience.

Remark 3. For a given value of k, the term within braces in (2.3.9) is $o(n^{-k-1})$; i.e., terms in the sum over k should tend to decrease as k increases, suggesting that the form given will provide a reasonable approximation.

2.4 Accuracy of asymptotic expansion.

In the previous section, we have derived an asymptotic expansion for the upper tail distribution of $\overline{S}_0$. However, the error made in truncating after a given number of terms has not yet been addressed. Further, the expansion is not uniform with respect to s. Thus the question of the accuracy of the approximation based on the asymptotic expansion naturally arises. Though, based on the results here, we have the exact distribution of $\overline{S}_0$ under only limited conditions, we can also investigate the general behavior of the asymptotic approximation. The following discussion is illustrated by selected values of the upper tail probability, computed from Theorem 2.2.2 and of the approximation provided by Theorem 2.3.1 with successively more terms (r = 0,1,2,3). Although Theorem 2.3.1 is valid only for $r \geq 1$, the case r = 0, i.e. $Pr(S>s) \doteq P(X>s-EY)$, is included here for comparison. The values of s have been selected to give similar sized tail probabilities. These results are displayed in Tables 2.4.1(a) - (d). Table 2.4.2 gives selected values of the approximate upper tail probability (using r = 3 in Theorem 2.3.1) for $\underline{\Omega} = \omega \underline{I}_p$ and other slightly different forms of $\underline{\Omega}$.

TABLE 2.4.1(a): Exact and Approximate Values of $Pr(\overline{S}_0 > s)$ for $p = 2$

				Approximate Probability			
n	ω	s	Exact Prob.	r = 0	r = 1	r = 2	r = 3
10	.25	6	.25850	.23817	.25640	.25884	.25901
		8	.10856	.09982	.10747	.10849	.10856
		10	.04551	.04184	.04504	.04547	.04550
		12	.01907	.01754	.01888	.01906	.01907
	.10	5	.29731	.28143	.29597	.29747	.29754
		7	.12184	.11524	.12120	.12181	.12184
		9	.04989	.04719	.04963	.04988	.04989
		12	.01307	.01236	.01300	.01307	.01307
	.05	5	.26895	.25657	.26791	.26896	.26901
		7	.10927	.10422	.10882	.10925	.10927
		9	.04439	.04233	.04420	.04438	.04439
		12	.01149	.01096	.01144	.01149	.01149
	.01	5	.24795	.23798	.24715	.24794	.24796
		7	.10007	.09604	.09974	.10006	.10007
		9	.04038	.03876	.04025	.04038	.04038
		12	.01035	.00994	.01032	.01035	.01035
20	.25	6	.24252	.23105	.24174	.24250	.24254
		8	.09567	.09114	.09536	.09566	.09567
		10	.03774	.03595	.03761	.03773	.03774
		12	.01489	.01418	.01484	.01488	.01489
	.10	5	.28610	.27720	.28564	.28610	.28611
		7	.11138	.10792	.11120	.11138	.11138
		9	.04336	.04201	.04329	.04336	.04336
		12	.01053	.01020	.01052	.01053	.01053
	.05	5	.25862	.25179	.25829	.25861	.25862
		7	.10023	.09759	.10011	.10023	.10023
		9	.03885	.03782	.03880	.03885	.03885
		12	.00937	.00913	.00936	.00937	.00937
	.01	5	.23845	.23300	.23821	.23845	.23845
		7	.09208	.08998	.09199	.09208	.09208
		9	.03556	.03475	.03552	.03556	.03556
		12	.00853	.00834	.00853	.00853	.00853

TABLE 2.4.1(a) (continued)

n	ω	s	Exact Prob.	Approximate Probability r = 0	r = 1	r = 2	r = 3
40	.25	6	.23320	.22720	.23299	.23320	.23321
		8	.08895	.08666	.08887	.08895	.08895
		10	.03393	.03305	.03390	.03393	.03393
		12	.01294	.01261	.01293	.01294	.01294
	.10	5	.27960	.27492	.27947	.27960	.27960
		7	.10590	.10413	.10585	.10590	.10590
		9	.04011	.03944	.04009	.04011	.04011
		12	.00935	.00919	.00934	.00935	.00935
	.05	5	.25281	.24925	.25273	.25282	.25282
		7	.09553	.09418	.09550	.09553	.09553
		9	.03610	.03559	.03608	.03610	.03610
		12	.00838	.00827	.00838	.00838	.00838
	.01	5	.23321	.23037	.23315	.23321	.23321
		7	.08795	.08688	.08793	.08795	.08795
		9	.03317	.03277	.03316	.03317	.03317
		12	.00768	.00759	.00768	.00768	.00768

TABLE 2.4.1(b) Exact and Approximate Values of $Pr(\overline{S}_0 > s)$ for p = 4

n	ω	s	Exact Prob.	Approximate Probability r = 0	r = 1	r = 2	r = 3
10	.25	12	.23599	.21950	.23573	.23629	.23605
		15	.08778	.07964	.08712	.08780	.08779
		17	.04343	.03899	.04298	.04342	.04343
		20	.01448	.01286	.01430	.01447	.01448
	.10	11	.21249	.20183	.21221	.21257	.21250
		13	.10772	.10118	.10734	.10774	.10773
		15	.05259	.04902	.05233	.05259	.05259
		19	.01166	.01076	.01158	.01166	.01166
	.05	10	.25505	.24492	.25494	.25513	.25506
		13	.09135	.08644	.09107	.09135	.09135
		15	.04396	.04135	.04378	.04396	.04396
		19	.00952	.00888	.00946	.00952	.00952
	.01	10	.22758	.21924	.22741	.22762	.22758
		12	.11436	.10922	.11411	.11436	.11436
		14	.05524	.05244	.05506	.05523	.05524
		18	.01196	.01127	.01190	.01195	.01196
20	.25	12	.21949	.20981	.21934	.21954	.21950
		14	.10817	.10235	.10792	.10817	.10817
		16	.05118	.04810	.05101	.05118	.05118
		20	.01061	.00989	.01056	.01061	.01061
	.10	10	.28134	.27401	.28134	.28137	.28135
		13	.09648	.09297	.09636	.09648	.09648
		15	.04486	.04305	.04478	.04486	.04486
		19	.00901	.00860	.00899	.00901	.00901
	.05	10	.24408	.23829	.24403	.24410	.24408
		12	.11917	.11564	.11907	.11917	.11917
		15	.05571	.05384	.05564	.05571	.05571
		18	.01123	.01080	.01121	.01123	.01123
	.01	10	.21716	.21246	.21711	.21717	.21717
		12	.10475	.10199	.10468	.10476	.10476
		14	.04852	.04709	.04847	.04852	.04852
		18	.00964	.00932	.00963	.00964	.00964

TABLE 2.4.1(b) (continued)

n	ω	s	Exact Prob.	Approximate Probability r = 0	r = 1	r = 2	r = 3
40	.25	12	.20982	.20461	.20978	.20984	.20983
		14	.10002	.09702	.09996	.10003	.10003
		16	.04578	.04424	.04573	.04578	.04578
		20	.00888	.00854	.00887	.00888	.00888
	.10	10	.27474	.27078	.27474	.27475	.27474
		13	.09058	.08877	.09054	.09058	.09058
		15	.04101	.04010	.04099	.04101	.04101
		18	.01190	.01161	.01189	.01190	.01190
	.05	10	.23784	.23475	.23784	.23786	.23785
		12	.11337	.11155	.11335	.11338	.11338
		14	.05172	.05078	.05070	.05072	.05072
		18	.00992	.00972	.00992	.00992	.00992
	.01	10	.21138	.20889	.21137	.21139	.21139
		12	.09972	.09830	.09970	.09973	.09973
		14	.04515	.04443	.04514	.04515	.04515
		18	.00857	.00841	.00856	.00857	.00857

TABLE 2.4.1(c): Exact and Approximate Values of $Pr(\overline{S}_0 > s)$ for $p = 6$

n	ω	s	Exact Prob.	Approximate Probability r = 0	r = 1	r = 2	r = 3
10	.25	17	.27453	.25784	.27518	.27483	.27453
		21	.09096	.08189	.09041	.09101	.09097
		23	.04932	.04373	.04885	.04933	.04933
		25	.02599	.02276	.02567	.02598	.02600
	.10	16	.21431	.20371	.21428	.21439	.21431
		20	.06519	.06039	.06491	.06519	.06519
		22	.03407	.03127	.03386	.03406	.03407
		24	.01735	.01580	.01722	.01734	.01735
	.05	15	.23594	.22646	.23597	.23600	.23594
		18	.09828	.09264	.09805	.09830	.09828
		20	.05207	.04865	.05188	.05208	.05208
		23	.01901	.01757	.01890	.01900	.01901
	.01	14	.26908	.26044	.26920	.26914	.26909
		17	.11319	.10774	.11301	.11321	.11320
		20	.04327	.04070	.04313	.04327	.04327
		22	.02195	.02053	.02186	.02195	.02195
20	.25	17	.25868	.24850	.25884	.25874	.25868
		20	.10594	.09978	.10575	.10595	.10594
		22	.05520	.05148	.05503	.05520	.05520
		25	.01955	.01803	.01946	.01955	.01955
	.10	15	.26791	.26106	.26799	.26793	.26791
		18	.10825	.10410	.10816	.10826	.10825
		20	.05581	.05332	.05573	.05581	.05581
		23	.01943	.01843	.01939	.01943	.01943
	.05	15	.22411	.21865	.22412	.22413	.22412
		18	.08781	.08478	.08774	.08781	.08781
		20	.04456	.04281	.04451	.04456	.04456
		22	.02192	.02100	.02188	.02192	.02192
	.01	14	.25899	.25399	.25903	.25901	.25900
		17	.10300	.10005	.10295	.10301	.10301
		19	.05258	.05084	.05253	.05258	.05258
		22	.01805	.01736	.01803	.01805	.01805

TABLE 4.2.1(c) (continued)

n	ω	s	Exact Prob.	Approximate Probability r = 0	r = 1	r = 2	r = 3
40	.25	17	.24905	.24345	.24910	.24907	.24907
		20	.09733	.09412	.09728	.09734	.09733
		22	.04910	.04724	.04905	.04910	.04910
		25	.01655	.01583	.01653	.01655	.01655
	.10	15	.26090	.25719	.26093	.26092	.26091
		18	.10162	.09946	.10160	.10163	.10162
		20	.05107	.04981	.05104	.05107	.05107
		23	.01710	.01661	.01708	.01710	.01710
	.05	15	.21744	.21452	.21745	.21746	.21746
		18	.08235	.08080	.08233	.08236	.08235
		20	.04082	.03944	.04080	.04082	.04082
		23	.01344	.01312	.01344	.01345	.01345
	.01	14	.25325	.25058	.25328	.25328	.25327
		17	.09765	.09613	.09764	.09766	.09766
		19	.04878	.04790	.04877	.04878	.04878
		22	.01619	.01586	.01619	.01619	.01619

TABLE 2.4.1(d): Exact and Approximate Values of $Pr(\overline{S}_0 > s)$ for $p = 8$

				Approximate Probability			
n	ω	s	Exact Prob.	r = 0	r = 1	r = 2	r = 3
10	.25	23	.24078	.22487	.24132	.24099	.24077
		26	.11584	.10459	.11544	.11593	.11584
		28	.06771	.06000	.06722	.06774	.06772
		30	.03836	.03344	.03794	.03836	.03836
	.10	20	.26630	.25514	.26665	.26638	.26630
		24	.09635	.08952	.09609	.09637	.09635
		26	.05460	.05012	.05435	.05461	.05460
		28	.03002	.02727	.02983	.03002	.03002
	.05	19	.27593	.26629	.27623	.27599	.27593
		23	.09921	.09323	.09901	.09922	.09921
		25	.05596	.05204	.05577	.05597	.05596
		27	.03060	.02820	.03045	.03060	.03060
	.01	19	.23504	.22683	.23514	.23508	.23504
		22	.10748	.10195	.10732	.10749	.10748
		25	.04496	.04207	.04481	.04496	.04496
		27	.02421	.02248	.02410	.02421	.02421
20	.25	23	.22347	.21384	.22360	.22351	.22347
		26	.09956	.09335	.09941	.09958	.09957
		28	.05511	.05110	.05495	.05512	.05511
		30	.02952	.02711	.02939	.02952	.02952
	.10	20	.25360	.24700	.25370	.25362	.25361
		23	.11314	.10864	.11307	.11315	.11315
		26	.04577	.04347	.04570	.04578	.04577
		28	.02403	.02269	.02398	.02403	.02403
	.05	19	.26477	.25909	.26486	.26479	.26478
		22	.11825	.11432	.11820	.11826	.11826
		25	.04778	.04576	.04772	.04778	.04778
		27	.02504	.02386	.02500	.02504	.02504
	.01	19	.22387	.21915	.22391	.22390	.22389
		22	.09707	.09408	.09703	.09708	.09708
		24	.05270	.05080	.05265	.05270	.05270
		27	.01983	.01899	.01980	.01983	.01983

TABLE 2.4.1(d) (continued)

n	ω	s	Exact Prob.	Approximate Probability r = 0	r = 1	r = 2	r = 3
40	.25	22	.27576	.27014	.27588	.27579	.27578
		26	.09083	.08760	.09079	.09084	.09084
		28	.04873	.04671	.04868	.04873	.04873
		30	.02528	.02411	.02524	.02528	.02528
	.10	20	.24622	.24262	.24626	.24624	.24623
		23	.10612	.10377	.10610	.10612	.10612
		25	.05715	.05566	.05713	.05715	.05715
		28	.02117	.02051	.02115	.02117	.02117
	.05	19	.25832	.25526	.25838	.25836	.25834
		22	.11182	.10977	.11182	.11184	.11184
		25	.04368	.04266	.04366	.04368	.04368
		27	.02236	.02178	.02235	.02236	.02236
	.01	18	.28245	.27977	.28250	.28248	.28248
		22	.09162	.09008	.09162	.09163	.09163
		24	.04871	.04775	.04870	.04872	.04872
		26	.02503	.02447	.02502	.02503	.02503

TABLE 2.4.2: Approximate Values of $Pr(\overline{S}_0 > s)$ - Using $r = 3$

p	n	s	ω						
			.10	.11	.12	.15	.05	.06	.075
p	n	s	.10	.09	.08	.05	.05	.04	.025
2	10	7	.12184	.12184	.12185	.12190	.10927	.10927	.10928
		9	.04989	.04989	.04990	.04992	.04439	.04439	.04439
		12	.01307	.01307	.01308	.01308	.01149	.01149	.01149
	20	7	.11138	.11138	.11139	.11140	.10023	.10023	.10024
		9	.04336	.04336	.04336	.04337	.03885	.03885	.03885
		12	.01053	.01053	.01053	.01054	.00937	.00937	.00937

p	n	s	ω								
			.10	.11	.11	.12	.15	.20	.05	.06	.075
			.10	.10	.11	.09	.10	.05	.05	.05	.050
p	n	s	.10	.09	.08	.09	.05	.05	.05	.04	.025
3	10	10	.11656	.11656	.11656	.11656	.11660	.11667	.10154	.10154	.10154
		12	.05314	.05314	.05315	.05315	.05316	.05320	.04575	.04575	.04576
		16	.01049	.01049	.01049	.01049	.01049	.01050	.00885	.00885	.00885
	20	10	.10550	.10550	.10550	.10550	.10552	.10555	.09206	.09206	.09206
		12	.04577	.04577	.04577	.04577	.04578	.04579	.03961	.03961	.03962
		16	.00817	.00817	.00817	.00817	.00817	.00818	.00698	.00698	.00698

p	n	s	ω								
			.10	.11	.12	.15	.15	.16	.05	.06	.08
			.10	.10	.11	.10	.125	.08	.05	.05	.06
			.10	.10	.09	.10	.075	.08	.05	.05	.04
p	n	s	.10	.09	.08	.05	.05	.08	.05	.04	.02
4	10	13	.10773	.10773	.10773	.10776	.10776	.10775	.09135	.09135	.09136
		15	.05259	.05260	.05260	.05261	.05262	.05261	.04396	.04396	.04397
		18	.01712	.01712	.01712	.01712	.01712	.01712	.01404	.01404	.01405
	20	13	.09648	.09648	.09648	.09649	.09649	.09649	.08184	.08184	.01814
		15	.04486	.04486	.04486	.04487	.04487	.04487	.03765	.03765	.03765
		18	.01356	.01356	.01356	.01356	.01356	.01356	.01123	.01123	.01123

(Computer programs used for the computations are available from the author.)

As expected, the accuracy of the approximation for a given number of terms improves as n increases. However, the approximation is quite good for n as small as 10 if s is sufficiently different from EV. For any values of p, n, and ω, the approximation worsens as s approaches EV. If s is very close to EV, the approximation can be extremely inaccurate, especially for odd p, for which higher derivatives of $h_s(y)$ become infinite. Obviously, the approximation should not be used for s close to EV. If $s \doteq EV$ or $s < EV$, other approximations may be possible. However, this question has not been addressed here. In other cases the approximation appears to be quite good, and the restriction to upper tail probabilities does not seem to be a serious drawback.

There appears to be some tendency for the approximation to worsen as p increases. However, this tendency is slight and does not appear to cause significant problems in using the expansion as an approximation. Finally, we note that, in most cases, a relatively accurate approximation can be obtained by including only terms of order n^{-1}; computations are simpler in that case, involving only the variance of a linear combination of chi-squared random variables and not the more complicated higher moments.

The figures in Table 2.4.2 indicate that the distribution of $\overline{S}_0$ is quite robust with respect to variability of $\underline{\Omega}$. The major factor in determining the distribution for a given value of p is the average of the ω's (or λ's). As would be expected, the tail probability increases as the sum of absolute deviations from $\overline{\omega}$ increases. However,

this effect is relatively unimportant when compared to the effect of $\overline{\omega}$, n and the difference between s and EV.

In Chapter 3, we shall investigate the noncentral distribution of $\overline{S}_0$ and derive some results under more complicated models. There too, additional extensions and unsolved questions are discussed.

Preceding Page Blank - Filmed

3. CLASSIFICATION UNDER MORE COMPLICATED MODELS

3.1 Noncentral distribution of $\overline{S}_0$.

When $S_{01}, S_{02}, \ldots, S_{0n}$ are as defined in paragraph (A) on pages 8-9, we have derived the null distribution of $\overline{S}_0 = n^{-1}\sum_{j=1}^{n} S_{0j}$; that is, when the distributions of $\underline{X}_0$ and $\underline{X}_1, \ldots, \underline{X}_n$ are identical. The resulting distribution reduces to that of a particular quadratic form in independent normal variables with zero means, or equivalently to a linear combination of independent central chi-squared variates.

For calculating the probabilities of misclassification in the case already studied, as well as in cases with more than one population to which the new object may be assigned, we are also interested in the distribution of $\overline{S}_0$ when the distribution of $\underline{X}_0$ is different from the common distribution of $\underline{X}_1, \ldots, \underline{X}_n$. When $E\underline{X}_0 = \underline{\mu}_0 \neq \underline{\mu} = E\underline{X}_i$ $(i=1,\ldots,n)$, using the same arguments as in the central case, we can clearly express $\overline{S}_0$ as the same quadratic form with the normal random variables having non-zero means, or in the equivalent form of a linear combination of independent noncentral chi-squared random variables. As we shall see in the following theorem, however, the actual representation obtained is somewhat simpler.

Theorem 3.1.1. Let $S_{01}, S_{02}, \ldots, S_{0n}$ be as defined in paragraph (A) on pages 8-9, with $\underline{\mu}_0 \neq \underline{\mu}$. Then $\overline{S}_0 = n^{-1}\sum_{j=1}^{n} S_{0j}$ can be represented as

$$\overline{S}_0 = \sum_{j=1}^{p} \{(1+n^{-1}\lambda_j)W'_{1j} + n^{-1}\lambda_j W_{2j}\}\ , \qquad (3.1.1)$$

where $\{W'_{1j}, W_{2j}\}_{j=1,\ldots,p}$ are mutually independent random variables, W'_{1j} being a noncentral chi-squared random variable with one degree of freedom and noncentrality parameter η_j^2 (η_j is defined in (3.1.3) in the proof), W_{2j} being a central chi-squared random variable with n-1 degrees of freedom, and $\lambda_1,\ldots,\lambda_p$ are the characteristic roots of $\underline{I}_p + 2\underline{\Sigma}^{-1}\underline{\Delta}$.

Proof: We can still write (as in the central case considered in Theorem 2.2.1) $\overline{S}_0 = n^{-1}\underline{U}'\underline{U}$ where

$$\underline{U}' = (\underline{U}'_{01}\underline{\Sigma}^{-\frac{1}{2}}, \underline{U}'_{02}\underline{\Sigma}^{-\frac{1}{2}}, \ldots, \underline{U}'_{0n}\underline{\Sigma}^{-\frac{1}{2}})$$

and $\underline{U}_{0j} = \underline{X}_0 - \underline{X}_j + \underline{Y}_{0j} - \underline{Y}_{j0}$. Thus $\underline{U}_{0j}$ is distributed as $N_p(\underline{\mu}_0-\underline{\mu}, 2(\underline{\Sigma}+\underline{\Delta}))$ and $\underline{U}$ is $N_{np}(\underline{\theta},\underline{V})$, where

$$\underline{\theta} = \begin{bmatrix} \underline{\Sigma}^{-\frac{1}{2}}(\underline{\mu}_0-\underline{\mu}) \\ \cdot \\ \cdot \\ \cdot \\ \underline{\Sigma}^{-\frac{1}{2}}(\underline{\mu}_0-\underline{\mu}) \end{bmatrix} \qquad \underline{V} = \begin{bmatrix} \underline{I}_p+2\underline{\Omega} & \underline{I}_p & \cdots & \underline{I}_p \\ \underline{I}_p & \underline{I}_p+2\underline{\Omega} & \cdots & \underline{I}_p \\ \vdots & \vdots & & \vdots \\ \underline{I}_p & \underline{I}_p & \cdots & \underline{I}_p+2\underline{\Omega} \end{bmatrix},$$

and $\underline{\Omega} = \underline{\Sigma}^{-\frac{1}{2}}\underline{\Delta}\underline{\Sigma}^{-\frac{1}{2}}$. Then, as before, by an appropriate transformation, we can express $\overline{S}_0$ as

$$\overline{S}_0 = n^{-1}\sum_{j=1}^{np} \alpha_j(Z_j+\eta_j)^2 , \qquad (3.1.2)$$

where $Z_1,\ldots,Z_{np}$ are independent unit normal random variables, η_j is the same function of the θ's as Z_j is of the U's, and $\alpha_1,\ldots,\alpha_{np}$ are the characteristic roots of $\underline{V}$. Since the α's are the same as in Theorem 2.2.1, all that remains to prove the result is to determine

the values of $\eta_1,\ldots,\eta_{np}$.

To find the η's, we must be more specific about the transformation leading to (3.1.2). Since $\underline{V}$ is symmetric we can write $\underline{V} = \underline{Q}\ \underline{D}\ \underline{Q}'$, where $\underline{D}$ is a diagonal matrix containing the characteristic roots of $\underline{V}$ and $\underline{Q}$ is the associated matrix of orthogonal characteristic vectors. Let $\underline{D}^{\frac{1}{2}}$ be the square root of $\underline{D}$ (i.e., $\underline{D}^{\frac{1}{2}}\underline{D}^{\frac{1}{2}} = \underline{D}$) and $\underline{D}^{-\frac{1}{2}}$ its inverse. Let $\underline{Z} = \underline{D}^{-\frac{1}{2}}\ \underline{Q}'(\underline{U}-\underline{\theta})$. Then $\underline{Z}$ is normal with mean $\underline{0}$ and covariance matrix $\underline{I}_{np}$. Since $\underline{U} = \underline{Q}\ \underline{D}^{\frac{1}{2}}\underline{Z} + \underline{\theta}$, if we let $\underline{\eta} = \underline{D}^{-\frac{1}{2}}\underline{Q}'\underline{\theta}$, we have $\underline{U} = \underline{Q}\ \underline{D}^{\frac{1}{2}}(\underline{Z}+\underline{\eta})$ and

$$\underline{U}'\underline{U} = (\underline{Z}+\underline{\eta})'\underline{D}(\underline{Z}+\underline{\eta}) = \sum_{j=1}^{np} \alpha_j (Z_j+\eta_j)^2 .$$

By Lemma 2.2.3, we know that p of the α's are equal to $n + \lambda_j$ $(j=1,\ldots,p)$, and the rest are repetitions of λ_j $(j=1,\ldots,p)$. Assume that we number the α's such that $\alpha_i = n + \lambda_i$ and $\alpha_{i+kp} = \lambda_i$ for $i = 1,2,\ldots,p$; $k = 1,2,\ldots,n-1$. Let $\underline{\Lambda}$ be the diagonal matrix containing $\lambda_1,\ldots,\lambda_p$ and $\underline{R}$ be the associated matrix of orthogonal characteristic vectors of $\underline{I}_p+2\underline{\Omega}$. Then by Lemma 2.2.3, we have

$$\underline{Q} = \begin{bmatrix} \frac{\underline{R}}{\sqrt{n}} & \frac{\underline{R}}{\sqrt{2}} & \frac{\underline{R}}{\sqrt{6}} & \cdots & \frac{\underline{R}}{\sqrt{n(n-1)}} \\ \frac{\underline{R}}{\sqrt{n}} & \frac{-\underline{R}}{\sqrt{2}} & \frac{\underline{R}}{\sqrt{6}} & \cdots & \frac{\underline{R}}{\sqrt{n(n-1)}} \\ \frac{\underline{R}}{\sqrt{n}} & \underline{0} & \frac{-2\underline{R}}{\sqrt{6}} & \cdots & \frac{\underline{R}}{\sqrt{n(n-1)}} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\underline{R}}{\sqrt{n}} & \underline{0} & \underline{0} & \cdots & \frac{-(n-1)\underline{R}}{\sqrt{n(n-1)}} \end{bmatrix} .$$

Thus

$$\underline{Q}'\underline{\theta} = \begin{bmatrix} \sqrt{n}\ \underline{R}\ \underline{\Sigma}^{-\frac{1}{2}}(\underline{\mu}_0-\underline{\mu}) \\ \underline{0} \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix} ,$$

and

$$\begin{pmatrix} \eta_1 \\ \vdots \\ \eta_p \end{pmatrix} = \underline{D}^*\underline{R}\ \underline{\Sigma}^{-\frac{1}{2}}(\underline{\mu}_0-\underline{\mu}) \ , \qquad \begin{pmatrix} \eta_{p+1} \\ \vdots \\ \eta_{np} \end{pmatrix} = \underline{0} \ , \tag{3.1.3}$$

where $\underline{D}^*$ is a diagonal matrix with elements $(1+n^{-1}\lambda_j)^{-\frac{1}{2}}$. The conclusion follows on substitution of (3.1.3) and the α's in (3.1.1). □

The exact distribution of $\overline{S}_0$ in the noncentral case is even more intractable than in the central case, even when the λ's are all equal, and a computational scheme will not be given here. However, an asymptotic expansion similar to that given in Theorem 2.3.1 for the central case can be obtained. It is given in the following theorem.

Theorem 3.1.2. Let $\overline{S}_0$, λ_j, η_j, W_{1j}', W_{2j} $(j = 1,2,\ldots,p)$ be as defined in Theorem 3.1.1. Then for $s > \sum_{j=1}^p \lambda_j$,

$$\Pr(\overline{S}_0 > s) = \sum_{k=0}^{r} d_k'\{\Pr(\chi^2_{p+2k} > (s-EV)/\beta)$$

$$+ \sum_{j=2}^{2r-2} \frac{a_{k,j}\mu_j}{j!}\} + o(n^{-r+1}) \ , \tag{3.1.4}$$

where $\beta = 1+n^{-1}\overline{\lambda}$, $\overline{\lambda} = p^{-1}\sum_{j=1}^p \lambda_j$, $V = n^{-1}\sum_{j=1}^p \lambda_j W_{2j}$, $\mu_j = E[(V-EV)/\beta]^j$, and

$$d_k' = (-1)^k \sum_{j=k}^{r} \binom{j}{k} c_j' \beta^{-j} \qquad (k=0,1,\ldots,r)$$

$$c_0' = 1$$

$$c_k' = (2k)^{-1} \sum_{j=0}^{k-1} B_{k-j}' c_j' \qquad (k=1,2,\ldots,r)$$

$$B_k' = \sum_{j=1}^{p} \left(\frac{\overline{\lambda}-\lambda_j}{n}\right)^k - k \sum_{j=1}^{p} \eta_j^2 (1+n^{-1}\lambda_j) \left(\frac{\overline{\lambda}-\lambda_j}{n}\right)^{k-1} \qquad (k=1,2,\ldots,r) \quad .$$

The coefficients $\{a_{k,j}\}_{k=0,1,\ldots;j=0,1,\ldots}$ are the same as in Theorem 2.3.1.

<u>Proof</u>: From Theorem 3.1.1, we have that $\overline{S}_0 = U+V$ where

$$U = \sum_{j=1}^{p} (1+n^{-1}\lambda_j) W_{1j}' \ , \qquad V = n^{-1} \sum_{j=1}^{p} \lambda_j W_{2j} \ .$$

But as in Theorem 2.3.1, for sufficiently large n, we can express the noncentral distribution of U using the Laguerre series expansion described in Johnson and Kotz [11], chapter 29, section 6.3. With the exception that the constants have a different form, the expansion is as in the central case in Theorem 2.3.1, i.e.,

$$\Pr(U>u) = \sum_{j=0}^{\infty} c_j' \beta^{-j} j! [\Gamma(\tfrac{1}{2}p)/\Gamma(\tfrac{1}{2}p+j)]$$

$$\times \int_u^{\infty} f_p(x\beta^{-1}) L_j^{(\frac{1}{2}p-1)}(\tfrac{1}{2}x\beta^{-1}) dx \ , \qquad (3.1.5)$$

where $\beta = 1+n^{-1}\overline{\lambda}$, $f_p(x)$ is the density of a χ_p^2 random variable, and

$$c_0' = 1$$

$$c_k' = (2k)^{-1} \sum_{j=0}^{k-1} B_{k-j}' c_j' \qquad (j \geq 1)$$

$$B'_k = \sum_{j=1}^{p}\left(\frac{\overline{\lambda}-\lambda_j}{n}\right)^k - k\sum_{j=1}^{p}\eta_j^2(1+n^{-1}\lambda_j)\left(\frac{\overline{\lambda}-\lambda_j}{n}\right)^{k-1} \quad (k \ge 1) .$$

Note that the only difference between (3.1.5) and (2.3.10) of Theorem 2.3.1 is the form of the B's. Using the definition of $L_r^{(\alpha)}(x)$ and reversing the order of summation we obtain

$$\Pr(U>u) = \sum_{k=0}^{r} d'_k \Pr(\chi^2_{p+2k} > u/\beta) + R_r(u) ,$$

where

$$R_r(u) = \sum_{j=r+1}^{\infty} c'_j\beta^{-j}\sum_{k=0}^{j}(-1)^k\binom{j}{k}\Pr(\chi^2_{p+2k} > u/\beta) .$$

In this case, for some M_1 (independent of n), we have

$$|B'_k| \le M_1\sum_{j=1}^{p}(1+n^{-1}\lambda_j)\left(\frac{\overline{\lambda}-\lambda_j}{n}\right)^{k-1} + \sum_{j=1}^{p}\left(\frac{\overline{\lambda}-\lambda_j}{n}\right)^k$$

$$= O(n^{-k+1}) ,$$

or for M_2 (independent of n), $|B'_k| \le M_2 n^{-k+1}$. Thus $|c'_k| \le \frac{1}{2}(M_2/n)^{k-1}$ and for sufficiently large n,

$$|R_r(u)| \le \frac{1}{2}\left(\frac{M_2}{n}\right)^r \sum_{j=0}^{\infty}\left(\frac{2M_2}{n\beta}\right)^j = o(n^{-r+1}) .$$

Therefore we have

$$\Pr(\overline{S}_0>s) = P(U+V>s)$$

$$= \sum_{k=0}^{r} d'_k \Pr(\chi^2_{p+2k} + V/\beta > s/\beta) + o(n^{-r+1}). \quad (3.1.6)$$

But V is as in Theorem 2.3.1, and hence V/β satisfies the conditions of Lemma 2.3.2. So by Lemma 2.3.1, for each k,

$$\Pr(\chi^2_{p+2k} + V/\beta > s/\beta) = \Pr(\chi^2_{p+2k} > (s-EV)/\beta)$$
$$+ \sum_{j=2}^{2r-2} \frac{a_{k,j}\mu_j}{j!} + o(n^{-r+1}) \ , \tag{3.1.7}$$

where the a's are the same as in Theorem 2.3.1. The final form (3.1.4) results from substitution of (3.1.7) in (3.1.6). □

It is possible that if terms are collected as in Theorem 2.3.1, those not included will be small, while those in a given grouping will be of a similar order of magnitude. However, due to the more complex nature of the coefficients, this conjecture has not been proved. We also remark here that if a computer program is available to compute the approximate central distribution in Theorem 2.3.1, it would be a relatively minor modification to use it for the noncentral distribution, the only major change being in the computation of the B's.

3.2 Classification of several new objects.

In most instances, classification procedures improve rapidly when several new similar objects are to be classified, rather than just one as in the cases we have investigated thus far. So it is natural to question whether the type of procedure based on imperfect distance data behaves in such a manner. A slightly different notation will have to be used in this case; however, the basic model is similar.

(B) For $k = 1,2$, let $\underline{X}_1^{(k)}, \underline{X}_2^{(k)}, \ldots, \underline{X}_{n_k}^{(k)}$ be independent and identically distributed random variables with the $N_p(\underline{\mu}_k, \underline{\Sigma})$ distribution, and let $\{Y_{ij}^{(1)}, \underline{Y}_{ji}^{(2)}\}_{i=1,\ldots,n_1; j=1,\ldots,n_2}$ be independent and identically distributed random variables with the $N_p(\underline{0}, \underline{\Delta})$ distribution. As before, we interpret the $\underline{X}$'s as the true (unobservable) positions of the objects in

question, and the $\underline{Y}$'s as errors made in determining those positions. Let

$$S_{ij} = (\underline{X}_i^{(1)} + \underline{Y}_{ij}^{(1)} - \underline{X}_j^{(2)} - \underline{Y}_{ji}^{(2)})' \underline{\Sigma}^{-1} (\underline{X}_i^{(1)} + \underline{Y}_{ij}^{(1)} - \underline{X}_j^{(2)} - \underline{Y}_{ji}^{(2)}).$$

Thus S_{ij} is the measured or perceived distance between object i in the new group and object j in the "known" group.

The natural extension of $\overline{S}_0$ to the model in paragraph (B) is then

$$\overline{S}_B = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} S_{ij} ,$$

the average of all distances between objects in group 1 and those in group 2. The following theorem gives the distribution of $\overline{S}_B$ under this model, with $\underline{\mu}_1 = \underline{\mu}_2$; that is, the null or central case when all objects are actually from the same population.

Theorem 3.2.1. Let $\{S_{ij}\}_{i=1,\ldots,n_1; j=1,\ldots,n_2}$ be as defined in paragraph (B) above, with $\underline{\mu}_1 = \underline{\mu}_2$. Let $\overline{S}_B = (n_1 n_2)^{-1} \sum_{j=1}^{n_1} \sum_{j=1}^{n_2} S_{ij}$. Then, for $n_1, n_2 \geq 2$, we can represent $\overline{S}_B$ as

$$\overline{S}_B = \sum_{j=1}^{p} \{(n_2^{-1} + n_1^{-1}\lambda_j) W_{1j} + n_1^{-1}\lambda_j W_{2j}$$

$$+ (n_2^{-1} + n_1^{-1}(\lambda_j - 1)) W_{3j} + n_1^{-1}(\lambda_j - 1) W_{4j}\} , \qquad (3.2.1)$$

where all of the W's are mutually independent random variables and for $j = 1,2,\ldots,p$, $W_{1j}, W_{2j}, W_{3j}, W_{4j}$ are distributed respectively as $\chi_1^2, \chi_{n_1-1}^2, \chi_{n_2-1}^2, \chi_{(n_1-1)(n_2-1)}^2$, and $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the characteristic roots of $\underline{I}_p + 2n_2^{-1}\underline{\Omega}$, $\underline{\Omega} = \underline{\Sigma}^{-\frac{1}{2}} \underline{\Delta}\, \underline{\Sigma}^{-\frac{1}{2}}$. (Note that the matrix of which the λ's are the characteristic roots is different than in the previous cases.)

Proof: Let $\underline{U}_{ij} = \underline{\Sigma}^{-\frac{1}{2}}(\underline{X}_i^{(1)} - \underline{X}_j^{(2)})$ and $\underline{T}_{ij} = \underline{\Sigma}^{-\frac{1}{2}}(\underline{Y}_{ij}^{(1)} - \underline{Y}_{ji}^{(2)})$. Then

$\overline{S}_B = (n_1 n_2)^{-1}\underline{W}'\underline{W}$, where

$$\underline{W}' = (\underline{W}_{11}',\ldots,\underline{W}_{1n_2}',\underline{W}_{21}',\ldots,\underline{W}_{2n_2}',\ldots,\underline{W}_{n_1 1}',\ldots,\underline{W}_{n_1 n_2}')$$

and $\underline{W}_{ij} = \underline{U}_{ij} + \underline{T}_{ij}$. But $\underline{W}$ is normally distributed with mean $\underline{0}$ and some covariance matrix $\underline{V}$. Thus, as in earlier proofs, we can represent $\overline{S}_B$ as

$$\overline{S}_B = (n_1 n_2)^{-1} \sum_{j=1}^{n_1 n_2 p} \alpha_j Z_j^2 ,$$

where $\alpha_1,\alpha_2,\ldots,\alpha_{n_1 n_2 p}$ are the characteristic roots of $\underline{V}$ and $Z_1,\ldots,Z_{n_1 n_2 p}$ are independent unit normal random variables. Thus we need to determine $\underline{V}$ and its characteristic roots.

First we note that $\underline{U}_{ij}$ is distributed $N_p(\underline{0},2\underline{I}_p)$, $\underline{T}_{ij}$ is distributed $N_p(\underline{0},2\underline{\Omega})$, and $\underline{U}_{ij}$ and $\underline{T}_{ij}$ are independent for any choice of i and j. The $\underline{T}$'s are mutually independent, but the $\underline{U}$'s are not. We can write $\underline{V}$ as a partitioned matrix with (i,j)-th element the $n_1 p \times n_1 p$ matrix which is the covariance of the i-th and j-th random $n_1 p$-vectors in $\underline{W}$. But

$$\begin{aligned}
\mathrm{cov}(\underline{W}_{ij},\underline{W}_{ij}) &= E\underline{W}_{ij}\underline{W}_{ij}' \\
&= E(\underline{U}_{ij}\underline{U}_{ij}' + \underline{T}_{ij}\underline{U}_{ij}' + \underline{U}_{ij}\underline{T}_{ij}' + \underline{T}_{ij}\underline{T}_{ij}') \\
&= \mathrm{cov}(\underline{U}_{ij},\underline{U}_{ij}) + \mathrm{cov}(\underline{T}_{ij},\underline{T}_{ij}) \\
&= 2\underline{I}_p + 2\underline{\Omega}.
\end{aligned}$$

Similarly, if $i \neq k$, $\mathrm{cov}(\underline{W}_{ij},\underline{W}_{kj}) = \underline{I}_p$; if $j \neq k$, $\mathrm{cov}(\underline{W}_{ij},\underline{W}_{ik}) = \underline{I}_p$; if $i \neq j$, $k \neq \ell$, $\mathrm{cov}(\underline{W}_{ik},\underline{W}_{j\ell}) = 0$. Thus

$$\underline{V} = \begin{bmatrix} \underline{V}_{11} & \cdots & \underline{V}_{1n_2} \\ \vdots & & \vdots \\ \underline{V}_{n_2 1} & \cdots & \underline{V}_{n_2 n_2} \end{bmatrix} ,$$

where

$$\underline{V}_{ii} = \begin{bmatrix} 2\underline{I}_p+2\underline{\Omega} & \underline{I}_p & \cdots & \underline{I}_p \\ \underline{I}_p & 2\underline{I}_p+2\underline{\Omega} & \cdots & \underline{I}_p \\ \vdots & \vdots & & \vdots \\ \underline{I}_p & \underline{I}_p & \cdots & 2\underline{I}_p+2\underline{\Omega} \end{bmatrix} \quad (n_1p\times n_1p) ,$$

and for $i \neq j$, $\underline{V}_{ij} = \underline{I}_{n_1p}$. We now wish to solve the determinantal equation $|\underline{V}-\alpha\underline{I}_{n_1n_2p}| = 0$. Using Lemma 2.2.1, with

$$\underline{A} = \begin{bmatrix} (1-\alpha)\underline{I}_p+2\underline{\Omega} & \underline{I}_p & \cdots & \underline{I}_p \\ \underline{I}_p & (1-\alpha)\underline{I}_p+2\underline{\Omega} & \cdots & \underline{I}_p \\ \vdots & \vdots & & \vdots \\ \underline{I}_p & \underline{I}_p & \cdots & (1-\alpha)\underline{I}_p+2\underline{\Omega} \end{bmatrix}$$

and $\underline{B} = \underline{I}_{n_1p}$, we obtain $|\underline{V}-\alpha\underline{I}_{n_1n_2p}| = |\underline{A}|^{n_2-1}|\underline{D}_1|$, where

$$\underline{D}_1 = \begin{bmatrix} (1-\alpha+n_2)\underline{I}_p+2\underline{\Omega} & \underline{I}_p & \cdots & \underline{I}_p \\ \underline{I}_p & (1-\alpha+n_2)\underline{I}_p+2\underline{\Omega} & \cdots & \underline{I}_p \\ \vdots & \vdots & & \vdots \\ \underline{I}_p & \underline{I}_p & \cdots & (1-\alpha+n_2)\underline{I}_p+2\underline{\Omega} \end{bmatrix} .$$

But the roots of $|A|^{n_2-1} = 0$ are just the characteristic roots of the matrix $\underline{D}_2 = \underline{I}_{n_1} \otimes \underline{C}_1 + \underline{J}_{n_1} \otimes \underline{I}_p$ where $\underline{C}_1 = 2\underline{\Omega}$, each occurring n_2-1 times. By Lemma 2.2.2, the characteristic roots of $\underline{D}_2$ are, for j = 1,...,p, $n_1 + 2\omega_j$, each occurring once, and $2\omega_j$, each occurring n_1-1 times, with $\omega_1,\ldots,\omega_p$ being the characteristic roots of $\underline{\Omega}$.

Similarly, the roots of $|D_1| = 0$ are the characteristic roots of

$D_3 = \underline{I}_{n_1} \otimes \underline{C}_2 + \underline{J}_{n_1} \otimes \underline{I}_p$, where $\underline{C}_2 = n_2\underline{I}_p + 2\Omega$, which by Lemma 2.2.2 consist of, for $j = 1,\ldots,p$, $n_1+n_2+2\omega_j$, each occurring once, and $n_2 + \omega_j$, each occuring n_1-1 times. Thus the set $\{\alpha_1,\alpha_2,\ldots,\alpha_{n_1 n_2 p}\}$ consists of, for $j = 1,\ldots,p$,

$n_1+n_2 + 2\omega_j$	each occurring once		
$n_2 + 2\omega_j$	"	"	n_1-1 times
$n_1 + 2\omega_j$	"	"	n_2-1 times
$2\omega_j$	"	"	$(n_1-1)(n_2-1)$ times .

But

$$\frac{n_1+n_2+2\omega_j}{n_1 n_2} = \frac{1}{n_2} + \frac{1+2n_2^{-1}\omega_j}{n_1} = \frac{1}{n_2} + \frac{\lambda_j}{n_1} ,$$

$$\frac{n_2+2\omega_j}{n_1 n_2} = \frac{1}{n_1}\left(1 + \frac{2\omega_j}{n_2}\right) = \frac{\lambda_j}{n_1} ,$$

$$\frac{n_1+2\omega_j}{n_1 n_2} = \frac{1}{n_2} + \frac{2n_2^{-1}\omega_j}{n_1} = \frac{1}{n_2} + \frac{\lambda_j - 1}{n_1} ,$$

$$\frac{2\omega_j}{n_1 n_2} = \frac{2n_2^{-1}\omega_j}{n_1} = \frac{\lambda_j - 1}{n_1} .$$

The result follows upon substitution, summing together the squared unit normal random variables having the same multipliers. □

With this characterization of $\overline{S}_B$, we can see that, for fixed n_1, the effect of the error in position determination becomes less important as n_2 gets larger, because $2n_2^{-1}\underline{\Omega}$ becomes smaller. Since the performance

of the classification rule depends in some measure on the variance of the classification criterion, we would like that variance to decrease as n_2 increases. In fact, the dominating term, $\Sigma(n_2^{-1}+n_1^{-1}\lambda_j)W_{ij}$, does have decreasing variance for increasing n_2. The third and fourth factors in the sum in (3.2.1) may offset some of that reduction; however, their contribution to the overall variance should tend to be small, due to the smallness of λ_j-1. The numerical example below tends to support this conclusion. Before giving the example, however, we note that the special case of $n_2 = 1$ can be derived from the form of (3.2.1), since the third and fourth factors in the sum are then degenerate (zero degrees of freedom) and putting $n_2 = 1$ in the first two factors does yield the characterization in Theorem 2.2.1.

<u>Example</u>: Let $\Omega = 0.1\ \underline{I}_2$, $n_1 = 10$. When $n_2 = 1$, $\mathrm{Var}(\overline{S}_0) = \mathrm{Var}(1.12\chi_2^2 + 0.12\chi_{18}^2) = 5.536$. When $n_2 = 2$, $\mathrm{Var}(\overline{S}_B) = \mathrm{Var}(0.61\chi_2^2 + 0.11\chi_{18}^2 + 0.51\chi_2^2 + 0.01\chi_{18}^2) \approx 2.968$. When $n_2 = 4$, $\mathrm{Var}(\overline{S}_B) = \mathrm{Var}(0.355\chi_2^2 + 0.105\chi_{18}^2 + 0.255\chi_6^2 + 0.005\chi_{54}^2) = 1.684$.

Thus, if we consider variance of the criterion as a measure of performance, we do appear to have better classification when there are more than one object available to be classified. We would also be interested then in computing probabilities of misclassification. The exact distribution of $\overline{S}_B$ will be rather complicated and obtaining a practical computational scheme is difficult. However, we can approximate the distribution using the asymptotic expansion developed earlier. The procedure will be outlined briefly here.

For n_2 relatively small and n_1 large, we can write $\overline{S}_B$ as

$\overline{S}_B = U+V$, where

$$U = \sum_{j=1}^{p} \{(n_2^{-1}+n_1^{-1}\lambda_j)W_{1j} + (n_2^{-1}+n_1^{-1}(\lambda_j-1))W_{3j}\},$$

$$V = \sum_{j=1}^{p} \{n_1^{-1}\lambda_j W_{2j} + n_1^{-1}(\lambda_j-1)W_{4j}\} , \tag{3.2.2}$$

with W_{1j}, W_{3j} independent χ_1^2 and $\chi_{n_1-1}^2$ random variables, and W_{2j}, W_{4j} independent $\chi_{n_1-1}^2$ and $\chi_{(n_1-1)(n_2-1)}^2$ random variables. Thus U is a quadratic form in normal variables with all multipliers of comparable size, and hence it will have a convergent Laguerre series expansion similar to that used in Theorem 2.3.1, the only changes being different formulas for β and the constants $\{B_k\}$.

It is also easy to verify that the k-th cumulant of V in (3.2.2) is of order $O(n^{1-r})$. Thus the conditions of Lemmas 2.3.1 and 2.3.2 are satisfied and we can derive an asymptotic expansion with the same form as (2.3.9), but with p replaced there by $n_2 p$. Everything else remains notationally the same, with

$$\beta = n_2^{-1} + n_1^{-1}\overline{\lambda} - n_1^{-1}n_2^{-1}(n_2-1)$$

$$\overline{\lambda} = p^{-1} \sum_{j=1}^{p} \lambda_j$$

$$B_k = \sum_{j=1}^{p} \left(\frac{\overline{\lambda}-\lambda_j}{n_1} - \frac{n_2-1}{n_1 n_2}\right)^k + (n_2-1)\sum_{j=1}^{p} \left(\frac{\overline{\lambda}-\lambda_j}{n_1} + \frac{1}{n_1 n_2}\right)^k ,$$

and $EV = (1 - n_1^{-1})[n_2\sum_{j=1}^{p}\lambda_j - p(n_2-1)]$ and μ_j is the j-th central moment of V/β, V being the random variable in (3.2.2). We saw in the case of $\overline{S}_0$ that the distribution was fairly insensitive to minor variations of the λ's. The distribution of $\overline{S}_B$ as given here should be even less sensitive to such variations, since the majority of the terms in the

primary expansion (of U in (3.2.2)) depend on λ_j only through $\lambda_j - 1$ which is nearly zero.

3.3 Discrimination between two populations.

If we consider the case of classifying a single new object as coming from one of two known populations, usually termed discrimination, we can achieve results similar to those obtained in the less complicated cases already investigated. Since the distribution involved is even more complicated than in the cases considered thus far, we will only give a representation for the criterion and make a conjecture concerning a way of approximating its distribution.

Using the notation of paragraph (B) on pages 53-54, $\underline{X}_1^{(k)},\ldots,\underline{X}_{n_k}^{(k)}$ are i.i.d. with $N_p(\underline{\mu}_k,\underline{\Sigma})$ distributions. Let $\underline{X}_0$ be an additional independent random variable distributed as $N_p(\underline{\mu}_0,\underline{\Sigma})$, and let $\underline{Y}_{0j}^{(i)}$, $\underline{Y}_{j0}^{(i)}$, $j = 1,2,\ldots,n_i$, $i = 1,2$ be independent random variables distributed as $N_p(\underline{0},\underline{\Delta})$. Let $\underline{S}_{0j}^{(i)} = (\underline{X}_0 + \underline{Y}_{0j}^{(i)} - \underline{X}_j^{(i)} - \underline{Y}_{j0}^{(i)})'\Sigma^{-1}(\underline{X}_0 + \underline{Y}_{0j}^{(i)} - \underline{X}_j^{(i)} - \underline{Y}_{j0}^{(i)})$, and $\overline{S}^{(i)} = n_i^{-1}\sum_{j=1}^{n_i} S_{0j}^{(i)}$.

If we first consider, as we did earlier, the case with the contaminating variables removed, we find that classification based on the difference $\overline{S}^{(1)} - \overline{S}^{(2)}$ is the natural analogue to what would be used in the absence of contamination. The following theorem gives a representation of the criterion $\overline{S}^{(1)} - \overline{S}^{(2)}$. As indicated following the proof, it may be possible to use an expansion for the distribution of an indefinite quadratic form to generalize Lemma 2.3.1 and give an approximation to this distribution.

Theorem 3.3.1. Let $\overline{S}^{(1)}$ and $\overline{S}^{(2)}$ be as described in the preceding discussion. Then the statistic $\overline{S}^{(1)} - \overline{S}^{(2)}$ can be represented as

$$\overline{S}^{(1)} - \overline{S}^{(2)} = \sum_{j=1}^{p}\{(1+n_1^{-1}\lambda_j)(W_{1j}^{(1)} + \eta_j^{(1)})^2 - (1+n_2^{-1}\lambda_j)(W_{1j}^{(2)} + \eta_j^{(2)})^2\}$$

$$+ \sum_{j=1}^{p}\{n_1^{-1}\lambda_j W_{1j}^{*} - n_2^{-1}\lambda_j W_{2j}^{*}\} \qquad (3.3.1)$$

where $\lambda_1,\ldots,\lambda_p$ are the characteristic roots of $\underline{I}_p + 2\underline{\Omega}$, $\underline{\Omega} = \underline{\Sigma}^{-\frac{1}{2}}\underline{\Delta}\,\underline{\Sigma}^{-\frac{1}{2}}$, $\{\underline{W}_{1j}, W_{1j}^{*}, W_{2j}^{*}\}_{j=1,\ldots,p}$ are mutually independent random variables, W_{ij}^{*} being chi-squared with n_i-1 degrees of freedom and

$$\underline{W}_{1j} = \begin{pmatrix} W_{1j}^{(1)} \\ W_{1j}^{(2)} \end{pmatrix}$$

being bivariate normal with mean $\underline{0}$ and covariance matrix

$$\begin{pmatrix} 1 & \gamma_j \\ \gamma_j & 1 \end{pmatrix},$$

with $\gamma_j = (n_1 n_2)^{\frac{1}{2}}[(n_1+\lambda_j)(n_2+\lambda_j)]^{-\frac{1}{2}}$ and $\eta_j^{(i)}$ is as defined in (3.3.3) in the proof.

Proof: Letting $\underline{U}_j^{(i)} = \underline{\Sigma}^{-\frac{1}{2}}(\underline{X}_0+\underline{Y}_{0j}^{(i)} - \underline{X}_j^{(i)}-\underline{Y}_{j0}^{(i)})$, $\underline{U}^{(i)\prime} = (\underline{U}_1^{(i)\prime},\ldots,\underline{U}_{n_i}^{(i)\prime})$, and $\underline{U}' = (\underline{U}^{(1)\prime}, \underline{U}^{(2)\prime})$, we have, as in several previous theorems, that $\underline{U}$ is normal with expected value $\underline{\theta}$ and covariance matrix $\underline{V}$, where

$$\underline{\theta} = \begin{pmatrix} \underline{\theta}^{(1)} \\ \underline{\theta}^{(2)} \end{pmatrix} \qquad \underline{V} = \begin{pmatrix} \underline{V}_{11} & \underline{V}_{12} \\ \underline{V}_{21} & \underline{V}_{22} \end{pmatrix}$$

$$\underline{\theta}^{(i)\prime} = [(\underline{\mu}_0-\underline{\mu}_i)'\underline{\Sigma}^{-\frac{1}{2}},\ldots,(\underline{\mu}_0\underline{\mu}_i)'\underline{\Sigma}^{-\frac{1}{2}}] \qquad (1\times n_i p)$$

$$\underline{V}_{ii} = \begin{bmatrix} 2\underline{I}_p+2\underline{\Omega} & \ldots & \underline{I}_p \\ \vdots & & \vdots \\ \underline{I}_p & \ldots & 2\underline{I}_p+2\underline{\Omega} \end{bmatrix} \qquad (n_i p\times n_i p)$$

$$\underline{V}_{ij} = \begin{bmatrix} \underline{I}_p & \cdots & \underline{I}_p \\ \vdots & & \vdots \\ \underline{I}_p & \cdots & \underline{I}_p \end{bmatrix} \quad (n_i p \times n_j p) \quad .$$

Let $\underline{\Lambda}$ be the diagonal matrix containing the characteristic roots of $\underline{I}_p + 2\underline{\Omega}$ and $\underline{R}$ be the associated matrix of orthogonal characteristic vectors; i.e., $\underline{I}_p + 2\underline{\Omega} = \underline{R}\,\underline{\Lambda}\,\underline{R}'$. Then by Lemma 2.2.3, we have $\underline{V}_{ii} = \underline{Q}_i \underline{D}_i \underline{Q}_i'$ where

$$\underline{Q}_i = \begin{bmatrix} \frac{\underline{R}}{\sqrt{n_i}} & \frac{\underline{R}}{\sqrt{2}} & \cdots & \frac{\underline{R}}{\sqrt{n_i(n_i-1)}} \\ \frac{\underline{R}}{\sqrt{n_i}} & \frac{-\underline{R}}{\sqrt{2}} & \cdots & \frac{\underline{R}}{\sqrt{n_i(n_i-1)}} \\ \vdots & \vdots & & \vdots \\ \frac{\underline{R}}{\sqrt{n_i}} & \underline{0} & \cdots & \frac{-(n_i-1)\underline{R}}{\sqrt{n_i(n_i-1)}} \end{bmatrix} \quad (n_i p \times n_i p)$$

and

$$\underline{D}_i = \begin{bmatrix} n_i \underline{I}_p + \underline{\Lambda} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{\Lambda} & \cdots & \underline{0} \\ \vdots & \vdots & & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{\Lambda} \end{bmatrix} \quad (n_i p \times n_i p) \quad .$$

Let

$$\underline{W} = \begin{pmatrix} \underline{W}^{(1)} \\ \underline{W}^{(2)} \end{pmatrix} = \begin{pmatrix} \underline{D}_1^{-\frac{1}{2}} \underline{Q}_1' & \underline{0} \\ \underline{0} & \underline{D}_2^{-\frac{1}{2}} \underline{Q}_2' \end{pmatrix} (\underline{U} - \underline{\theta})$$

with the partitioning and indexing of $\underline{W}$ the same as that of $\underline{U}$. Then $\underline{W}$ is normal with expected value $\underline{0}$ and covariance matrix

$$\underline{V}^* = \begin{pmatrix} V^*_{11} & \underline{V}^*_{12} \\ V^*_{21} & \underline{V}^*_{22} \end{pmatrix} ,$$

where $\underline{V}^*_{ij} = \underline{D}_i^{-\frac{1}{2}}\underline{Q}_i'\underline{V}_{ij}\underline{Q}_j\underline{D}_j^{-\frac{1}{2}}$. Thus $\underline{V}^*_{ii} = \underline{I}_{n_i p}$ and for $i \neq j$

$$V^*_{ij} = \begin{pmatrix} \underline{\Gamma} & \underline{O} \\ \underline{O} & \underline{O} \end{pmatrix} \qquad (n_i p \times n_j p) ,$$

where

$$\underline{\Gamma} = \begin{pmatrix} \gamma_1 & & 0 \\ & \ddots & \\ 0 & & \gamma_p \end{pmatrix}$$

and

$$\gamma_j = \left\{ \frac{n_1 n_2}{(n_1+\lambda_j)(n_2+\lambda_j)} \right\}^{\frac{1}{2}} .$$

Hence

$$\Pr(\overline{S}^{(1)} - \overline{S}^{(2)} \le y) = \Pr(n_1^{-1}\underline{U}^{(1)\prime}\underline{U}^{(1)} - n_2^{-1}\underline{U}^{(2)\prime}\underline{U}^{(2)} \le y)$$

$$= \Pr[n_1^{-1}(\underline{W}^{(1)} + \underline{D}_1^{-\frac{1}{2}}\underline{Q}_1'\underline{\theta}^{(1)})'\underline{D}_1(\underline{W}^{(1)} + \underline{D}_1^{-\frac{1}{2}}\underline{Q}_1'\underline{\theta}^{(1)})$$

$$- n_2^{-1}(\underline{W}^{(2)} + \underline{D}_2^{-\frac{1}{2}}\underline{Q}_2'\underline{\theta}^{(2)})'\underline{D}_2(\underline{W}^{(2)} + \underline{D}_2^{-\frac{1}{2}}\underline{Q}_2'\underline{\theta}^{(2)}) \le y] . \qquad (3.2.2)$$

As in Theorem 3.1.1,

$$\underline{Q}_i'\underline{\theta}^{(i)} = \begin{pmatrix} n_i^{\frac{1}{2}}\underline{R}\ \underline{\Sigma}^{-\frac{1}{2}}(\underline{\mu}_0-\underline{\mu}_i) \\ 0 \\ \vdots \\ 0 \end{pmatrix} ,$$

and among the elements of $\underline{W}$, the only nonzero covariances are $\mathrm{cov}(W_{1j}^{(1)}, W_{1j}^{(2)}) = \gamma_j$ for $j = 1,\ldots,p$. Thus if we let

$$\eta^{(i)} = \begin{bmatrix} \eta_1^{(i)} \\ \vdots \\ \eta_p^{(i)} \end{bmatrix} = (\underline{I}_p + n_i^{-1}\underline{\Lambda})^{-\frac{1}{2}}\underline{R}\ \underline{\Sigma}^{-\frac{1}{2}}(\underline{\mu}_0-\underline{\mu}_1) \qquad (3.3.3)$$

and

$$W^*_{ij} = \sum_{k=2}^{n_i} (W_{jk}^{(i)})^2 \quad (i = 1,2;\ j = 1,2,\ldots,p) ,$$

and substitute these quantities in (3.3.2), we obtain the desired representation. □

Of course, it is possible to make an additional transformation to obtain a representation eliminating the remaining nonzero covariances, by using the fact that

$$\begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix} = \tfrac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1+\gamma & 0 \\ 0 & 1-\gamma \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} .$$

For computational purposes this would be done; however, this final transformation leads to a notationally less convenient form.

Although development of a computational scheme or approximation has not been attempted here, the similarity of the representation (3.3.1) to (3.1.1) and (2.2.3) suggests that an asymptotic expansion might be feasible. There is still a dominant term (which in this case tends to the difference between two noncentral chi-squared random variables) and an additional term which is tending to a constant (in this case zero) in probability. The dominant term is an indefinite quadratic form and so the theory developed for the previous cases is not applicable. However, Press [20] gives a series expansion for the density of such a form which is similar to those used here for the positive definite forms. Thus it may be possible to generalize the results given here to that case or to prove similar results for the more general case.

3.4 Extensions and unsolved problems.

Several possible extensions of the results developed in Sections 2 and 3 have already been indicated, together with conjectures as to how they might be approached. We wish to indicate here some other interesting and desirable generalizations.

One of the restrictions placed on the models under consideration in Chapters 2 and 3 was that the covariance matrices of the distributions of both the unobservable true distances and of the error components be known (in the central cases in Chapter 2, only $\underline{\Omega} = \underline{\Sigma}^{-\frac{1}{2}} \underline{\Delta}\, \underline{\Sigma}^{-\frac{1}{2}}$ was needed). Certainly in many situations that assumption will not be satisfied. In the central cases, if we assume that $\underline{\Delta}$ is a constant multiple of $\underline{\Sigma}$ with the constant unknown, we would like to develop a classification procedure and study the distributions involved when the constant is estimated. Since, in this case, the expected value of the observed distances is a function of the unknown constant, it might be possible to construct an estimate based on the observed data and to obtain some useful results. Estimation of more general covariance structures, however, appears less promising. In that case we would require information about the individual characteristic roots of $\underline{\Omega}$, while the expected value of the observed distances depends on the trace of $\underline{\Omega}$, but not otherwise on the individual roots.

Of course, it would be desirable to eliminate as many of the assumptions on the distribution of the distances as possible and to analyze the problem with a nonparametric approach. Some attempts were made in this direction with little success. The primary difficulty is that, under any reasonable model, the observed distances cannot be

considered to be independent observations. Perhaps with some mild assumptions concerning the nature of the dependence, some useful nonparametric results could be obtained.

In a nonparametric approach, presumably no assumptions would be made concerning the dimensionality of the underlying model. The related problem of determination of the dimensionality based only on distance measurements among the objects would be another problem of interest.

Finally, application of the results obtained here to cluster analysis or scaling problems would be desirable. These applications would introduce more complexity because of their sequential nature. As was mentioned in the introduction, these problems motivated the original inquiry into analysis based on distances. If applications of these results in those areas could be developed, it would be gratifying.

4. NONCENTRALITY ESTIMATION

4.1 Introduction.

In dealing with problems involving distance measurements, one often encounters the noncentral chi-squared distribution. We have seen several such instances in the preceding chapters. An even simpler example in which the distribution arises is the following:

Let P_1 and P_2 be points in euclidean two-space with coordinates (x_{11},x_{12}) and (x_{21},x_{22}) respectively. For $i,j = 1,2$, let the error made in determining the coordinate x_{ij} be ε_{ij}, where the errors are independent and identically distributed $N(0,\sigma^2)$ random variables. Putting $y_{ij} = x_{ij} + \varepsilon_{ij}$, the measured squared euclidean distance between P_1 and P_2, $\sum_{i=1}^{2}(y_{1i}-y_{2i})^2$, is distributed as $2\sigma^2$ times a noncentral chi-squared random variable with two degrees of freedom and noncentrality parameter $(2\sigma^2)^{-1}$ times the true squared distance between P_1 and P_2, $\sum_{i=1}^{2}(x_{1i}-x_{2i})^2$.

If we are interested, then, in making inferences concerning the true distances in such a situation involving measurement error or other similar types of errors, we are naturally led to the problem of estimation of the noncentrality parameter of a noncentral chi-squared random variable. A further complication of the problem can be introduced if we assume that our observations are not actually measurements of the true distances, but rather some monotonic transformation of those distances. Such an assumption is made, for example, in the multi-

dimensional scaling analysis mentioned in Chapter 1. In this chapter, we shall introduce an estimation procedure based on the two-sample Wilcoxon-Mann-Whitney statistic, which can be used when the observed data are the result of such a monotonic transformation. We will investigate the properties of the resulting estimator and compare them with those of the maximum likelihood estimation which would be applicable if the data were not transformed.

We shall need two representations of the density function of $\chi_p'^2(\lambda)$, where $\chi_p'^2(\lambda)$ stands for a noncentral chi-squared random variable with p degrees of freedom and noncentrality parameter λ. (If $\lambda = 0$, we will continue to use χ_p^2 instead of $\chi_p'^2(0)$.) The first representation is

$$f_{p,\lambda}(u) = \sum_{j=0}^{\infty} (\lambda/2)^j (j!)^{-1} \exp(-\lambda/2) f_{p+2j}(u) \quad (u \geq 0), \qquad (4.1.1)$$

where

$$f_p(u) = \tfrac{1}{2}(\tfrac{1}{2}u)^{\frac{1}{2}p-1} [\Gamma(\tfrac{1}{2}p)]^{-1} \exp(-u/2) \quad (u \geq 0) \qquad (4.1.2)$$

is the central chi-squared density. The second representation is

$$f_{p,\lambda}(u) = \tfrac{1}{2}(u/\lambda)^{\frac{1}{4}(p-2)} \exp[-\tfrac{1}{2}(\lambda+u)] I_{\frac{1}{2}p-1}(\sqrt{\lambda u}) \quad (u \geq 0), \qquad (4.1.3)$$

where $I_p(x)$ is the modified Bessel function of the first kind of order p. These formulas are given, among other places, in Johnson and Kotz [11], chapter 28.

4.2 Maximum likelihood estimation.

In the situation in which the observed data are the actual measurements and not a transformation of them, maximum likelihood estimation would be applicable and the resulting estimator will have desirable properties. Although the estimator to be proposed will apply in the

more general case with transformed measurements, it will be of interest to compare it with the maximum likelihood estimator in the restricted case. Thus we first wish to discuss some results of the maximum likelihood approach.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with density function (4.1.3) where p is known and λ is unknown. Letting $q = \frac{1}{2}p-1$ and denoting the likelihood function by L, we have

$$L = 2^{-n}\lambda^{-\frac{1}{2}nq} \prod_{j=1}^{n} \{x_j^{\frac{1}{2}q} I_q(\sqrt{\lambda x_j})\} \exp\{-\tfrac{1}{2}\sum_{j=1}^{n}(\lambda + x_j)\} \quad .$$

Thus

$$\log L = -n \log 2 - \tfrac{1}{2}nq \log \lambda$$
$$+ \sum_{j=1}^{n} \{\tfrac{1}{2}q \log x_j + \log I_q(\sqrt{\lambda x_j})\} - \tfrac{1}{2}\sum_{j=1}^{n}(\lambda + x_j) \quad . \tag{4.2.1}$$

Since

$$\frac{\partial}{\partial\lambda} I_q(\sqrt{\lambda u}) = [I_{q+1}(\sqrt{\lambda u}) + \frac{q}{\sqrt{\lambda u}} I_q(\sqrt{\lambda u})] \frac{\partial}{\partial\lambda}(\sqrt{\lambda u})$$
$$= \tfrac{1}{2}(u/\lambda)^{\frac{1}{2}} I_{q+1}(\sqrt{\lambda u}) + \tfrac{1}{2}(q/\lambda) I_q(\sqrt{\lambda u}) \quad , \tag{4.2.2}$$

we have from 4.2.1, upon differentiating, the likelihood equation

$$n\sqrt{\lambda} = \sum_{j=1}^{n} \frac{I_{q+1}(\sqrt{\lambda x_j})}{I_q(\sqrt{\lambda x_j})} \sqrt{x_j} \quad . \tag{4.2.3}$$

Thus, if (4.2.3) has a solution, it will provide us with the maximum likelihood estimator (MLE). The above derivation has been given by Meyer [18] for the special case of p = 2 and by Pandey and Rahman [19] in a modified form for the general case; it is given here for completeness. Both Meyer and Pandey and Rahman also give the following results

(again for the case p = 2 in Meyer):

(a) if $\sum_{j=1}^{n} X_j > np$, then a unique solution of (4.2.3) exists and gives the MLE of λ_j;

(b) if $\sum_{j=1}^{n} X_j \leq np$, the MLE of λ is 0;

(c) $\lim_{n\to\infty} \Pr(\sum_{j=1}^{n} X_j > np) = 1$; that is, as $n \to \infty$, the probability approaches one that the MLE is based on the observations through equation (4.2.3).

In investigating the efficiency of the alternative estimator to be proposed, we need the asymptotic distribution of the MLE of λ. It is well known that under certain regularity conditions, which can be shown to be satisfied in this case, the MLE is asymptotically normal, unbiased, and efficient, i.e., attains the Cramér-Rao bound. The following theorem states these results more precisely, without proof, and gives the form of the asymptotic variance.

Theorem 4.2.1. Let λ^* be the MLE, based on a random sample of n observations, of the noncentrality parameter of a $\chi_p'^2(\lambda)$ random variable. Then $\sqrt{n}(\lambda^*-\lambda)$ is asymptotically normally distributed with mean zero and variance $V_p^*(\lambda)$, where (with $q = \frac{1}{2}p-1$)

$$[V_p^*(\lambda)]^{-1} = -\frac{1}{4} + \frac{1}{8} e^{-\frac{1}{2}\lambda} \int_0^\infty \left(\frac{u}{\lambda}\right)^{\frac{1}{2}q+1} \frac{I_{q+1}^2(\sqrt{\lambda u})}{I_q(\sqrt{\lambda u})} e^{-\frac{1}{2}u} du .$$

Proof: As mentioned prior to the statement of the theorem, it is well known that under the conditions of this theorem, $\sqrt{n}(\lambda^*-\lambda)$ is asymptotically normally distributed with mean zero and variance

$$[E\{\frac{\partial}{\partial\lambda} \log f_{p,\lambda}(u)\}^2]^{-1}$$

(see, e.g., Kendall and Stuart [14], section 18.16). Thus we have only to show that

$$[V_p^*(\lambda)]^{-1} = E\{\frac{\partial}{\partial\lambda} \log f_{p,\lambda}(u)\}^2 .$$

From (4.1.3) and (4.2.2) we have

$$\frac{\partial}{\partial\lambda} \log f_{p,\lambda}(u) = -\tfrac{1}{2} + \tfrac{1}{2}\left(\frac{u}{\lambda}\right)^{\frac{1}{2}} \frac{I_{q+1}(\sqrt{\lambda u})}{I_q(\sqrt{\lambda u})} .$$

Thus

$$E\{\frac{\partial}{\partial\lambda} \log f_{p,\lambda}(u)\}^2$$

$$= \tfrac{1}{4} - \tfrac{1}{2}E\left\{\left(\frac{u}{\lambda}\right)^{\frac{1}{2}} \frac{I_{q+1}(\sqrt{\lambda u})}{I_q(\sqrt{\lambda u})}\right\} + \tfrac{1}{4}E\left\{\left(\frac{u}{\lambda}\right) \frac{I^2_{q+1}(\sqrt{\lambda u})}{I^2_q(\sqrt{\lambda u})}\right\} , \tag{4.2.4}$$

where the expectation is taken with respect to $f_{p,\lambda}(u)$. But

$$\left(\frac{u}{\lambda}\right)^{\frac{1}{2}} \frac{I_{q+1}(\sqrt{\lambda u})}{I_q(\sqrt{\lambda u})} f_{p,\lambda}(u) = f_{p+2,\lambda}(u) ,$$

and

$$\left(\frac{u}{\lambda}\right) \frac{I^2_{q+1}(\sqrt{\lambda u})}{I^2_q(\sqrt{\lambda u})} f_{p,\lambda}(u) = \tfrac{1}{2}\left(\frac{u}{\lambda}\right)^{\frac{1}{2}q+1} \frac{I^2_{q+1}(\sqrt{\lambda u})}{I_q(\sqrt{\lambda u})} e^{-\frac{1}{2}(\lambda+u)} .$$

The result follows on substitution in (4.2.4). □

4.3 <u>A randomized estimation procedure</u>.

While the maximum likelihood estimator has many desirable properties, an explicit solution to (4.2.3) does not exist. Of course, since $I_q(x)$ is a known, tabulated function, the equation can be solved numerically. However, from the computational standpoint an estimation which is an explicit function of the sample values is often preferable. In addition, we would like to be able to estimate λ even if the

observations have been subjected to a monotone transformation. In this section, we will develop an alternative estimation procedure, based on the two-sample Wilcoxon-Mann-Whitney statistic, which satisfies these requirements.

Before introducing the proposed estimator, we shall give some general results on the Wilcoxon-Mann-Whitney statistic, which we will use later. (Because of varying ways of identifying the samples and sample sizes, several of the results given here appear slightly different than they do in the cited sources.)

Theorem 4.3.1. Let $X_1,\ldots,X_n$; $Y_1,\ldots,Y_m$ be independent sets of i.i.d. random variables, the distributions of X_i and Y_j being (possibly) different. Let

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} c(Y_i - X_j)$$

where $c(x) = 1$ or 0 according as $x > 0$ or $x \le 0$. Then the following hold:

(a) U is an unbiased estimator of $mn\theta$, where $\theta = \Pr(X_1 < Y_1)$;

(b) if $m/n \to k$, a positive constant, as $m,n \to \infty$, then U/mn is asymptotically normally distributed;

(c) $\mathrm{Var}(U) = mn[(m-1)\phi^2 + (n-1)\gamma^2 + \theta(1-\theta)]$, where

$$\phi^2 = \Pr(X_1 < Y_1,\ X_1 < Y_2) - \theta^2$$

$$\gamma^2 = \Pr(X_1 < Y_1,\ X_2 < Y_1) - \theta^2 ;$$

(d) $\mathrm{Var}(U) \le mn\theta(1-\theta)\max(m,n)$.

Proof: (a) and (b) follow directly from the theory of U-statistics (see, e.g., Fraser [6]); (c) and (d) are given by Van Dantzig [26].

□

We return now to the specific problem of estimating the noncentrality parameter of a noncentral chi-squared random variable. Suppose that $X_1,\ldots,X_n$ are i.i.d. random variables with density (4.1.1); i.e., X_i is distributed as $\chi_p'^2(\lambda)$. Suppose further that $Y_1,\ldots,Y_m$ are i.i.d. random variables with density (4.1.2); i.e., Y_i is distributed as χ_p^2. In practice, only $X_1,X_2,\ldots,X_n$ may be available as observed data. In that case, $Y_1,Y_2,\ldots,Y_m$ could be generated numerically or taken from a table of random deviates. Since a χ_p^2 random variable is the sum of p independent squared unit normal random variables, widely available tables of random normal deviates could be used to generate the Y_i's if necessary. Hence the terminology randomized estimator. Alternatively, both the X_i's and Y_i's may be observed data, the Y_i's being thought of as "control estimates" of the noncentrality parameter zero. In this case, both the X_i's and Y_i's could be assumed to have been transformed by the same monotonic transformation without affecting the remaining analysis. If the Y_i's are generated randomly, m can be chosen for convenience to be an exact multiple of n, say m = rn; while not necessary for the remaining analysis, such a choice simplifies several of the results and will be assumed here. Let

$$t = (mn)^{-1}U \,, \tag{4.3.1}$$

where U is as defined in Theorem 4.3.1. It follows immediately from (a) of that theorem that t is an unbiased estimator of $\theta = \Pr(X_1<Y_1)$ and from (d) that it is consistent for θ as $n \to \infty$. In order to derive an estimator for the unknown parameter λ, it will be sufficient to express λ as a known function of θ. We now proceed to determine that function.

Using (4.1.1) and (4.1.2), we can express θ as

$$\theta = \Pr(X_1 < Y_1) = \int_{y=0}^{\infty} \int_{x=0}^{y} f_{p,\lambda}(x) f_p(y)\,dx\,dy$$

$$= \int_{y=0}^{\infty} \int_{x=0}^{y} \sum_{j=0}^{\infty} (\tfrac{1}{2}\lambda)^j (j!)^{-1} \exp(-\tfrac{1}{2}\lambda) f_{p+2j}(x) f_p(y)\,dx\,dy\ .$$

Since all terms are positive and $0 \le \theta \le \tfrac{1}{2}$, we may interchange the summation and integration, obtaining

$$\theta = \sum_{j=0}^{\infty} (\tfrac{1}{2}\lambda)^j (j!)^{-1} \exp(-\tfrac{1}{2}\lambda) \int_{y=0}^{\infty} \int_{x=0}^{y} f_{p+2j}(x) f_p(y)\,dx\,dy\ .$$

But each integral in the above equation is a function of p and j only, $a_{p,j}$ say, and in fact can be expressed as an incomplete beta ratio:

$$a_{p,j} = \int_{y=0}^{\infty} \int_{x=0}^{y} f_{p+2j}(x) f_p(y)\,dx\,dy = I_{\frac{1}{2}}(\tfrac{1}{2}p+j, \tfrac{1}{2}p)\ ,$$

where $I_x(p,q)$ is the incomplete beta ratio. Letting

$$g_p(\lambda) = \sum_{j=0}^{\infty} a_{p,j} (\tfrac{1}{2}\lambda)^j (j!)^{-1} \exp(-\tfrac{1}{2}\lambda)\ , \qquad (4.3.2)$$

we can now express θ as an explicit, known function of λ: $\theta = g_p(\lambda)$. However, we require λ as a function of θ. We now proceed to investigate the properties of $g_p(\lambda)$ and show that a single-valued inverse function exists.

To show that such an inverse function exists, it will be sufficient to show that the first derivative of $g_p(\lambda)$ is negative for all λ. From (4.3.2) we have

$$g_p'(\lambda) = \frac{d}{d\lambda}\{e^{-\frac{1}{2}\lambda}\sum_{j=0}^{\infty} a_{p,j}(j!)^{-1}(\tfrac{1}{2}\lambda)^j\}$$

$$= \tfrac{1}{2}e^{-\frac{1}{2}\lambda}\sum_{j=0}^{\infty}(\tfrac{1}{2}\lambda)^j(j!)^{-1}(a_{p,j+1} - a_{p,j}) \ . \qquad (4.3.3)$$

Since $a_{p,j} = I_{\frac{1}{2}}(\frac{1}{2}p+j,\ \frac{1}{2}p)$, it follows that

$$a_{p,j} - a_{p,j+1} = \frac{\Gamma(p+j)}{\Gamma(\frac{1}{2}p+j+1)\Gamma(\frac{1}{2}p)}(\tfrac{1}{2})^{p+j} > 0, \qquad (4.3.4)$$

and therefore $g_p'(\lambda) < 0$ for all λ. Thus for $p \geq 2$, $g_p(\lambda)$ is monotonically decreasing and therefore $g_p^{-1}(\theta)$ exists.

For even values of p, the function $g_p(\lambda)$ has a closed form, which can be derived using, once again, the recurrence properties of the incomplete beta ratio. In particular, we have

$$g_2(\lambda) = \tfrac{1}{2}\exp(-\tfrac{1}{4}\lambda)$$

$$g_4(\lambda) = \tfrac{1}{2}\exp(-\tfrac{1}{4}\lambda)\{1 + \frac{\lambda}{16}\}$$

$$g_6(\lambda) = \tfrac{1}{2}\exp(-\tfrac{1}{4}\lambda)\{1 + \frac{3\lambda}{32} + \frac{\lambda^2}{512}\} \ .$$

In general, if $q = \frac{1}{2}p-1$ is a non-negative integer, then $g_p(\lambda) = g_2(\lambda)P_q(\lambda)$, where $P_q(\lambda)$ is a polynomial of degree q in λ such that $P_q(0) = 1$. While $g_p^{-1}(\cdot)$ has a closed form only for p = 2, it is a simple matter to construct tables of $g_p^{-1}(\cdot)$ for any value of p. Such tables of $g_p^{-1}(\cdot)$ are included in the appendix for several values of p.

Having $g_p^{-1}(\theta)$ either in explicit or tabular form, we are now in a position to define our randomized estimator of the unknown noncentrality parameter λ.

Theorem 4.3.2. Let t be as defined in (4.3.1) and $g_p(\lambda)$ as defined in (4.3.2). Then $\hat{\lambda} = g_p^{-1}(t)$ is an estimator of λ with the following properties:

(a) $\hat{\lambda}$ is consistent for λ as $n \to \infty$;

(b) $\hat{\lambda}$ is asymptotically unbiased;

(c) $\sqrt{n}\,\hat{\lambda}$ is asymptotically normally distributed with variance

$$V_p(\lambda) = [\frac{d}{d\theta} g_p^{-1}(\theta)]^2[\phi^2 + r^{-1}\gamma^2] ,$$

where ϕ^2 and γ^2 are as defined in Theorem 4.3.1 and $r = m/n$.

Proof: All three statements will follow if we can show that $\sqrt{n}(\hat{\lambda}-\lambda)$ has an asymptotic normal distribution with mean zero and variance as given in (c). Since we assume $r = m/n$ is a constant, it follows from (a) and (b) of Theorem 4.3.1 that $\sqrt{n}(t-\theta)$ is asymptotically normally distributed with zero mean. From (c) of that theorem, we have

$$\begin{aligned} \mathrm{Var}(\sqrt{n}t) &= \mathrm{Var}[(m\sqrt{n})^{-1}U] = (r^{-2}n^{-3})\mathrm{Var}(U) \\ &= \phi^2 + r^{-1}\gamma^2 + (rn)^{-1}[\phi^2+\gamma^2+\theta(1-\theta)] . \end{aligned}$$

Hence $\lim_{n\to\infty}[\mathrm{Var}(\sqrt{n}t)] = \phi^2 + r^{-1}\gamma^2$, and it follows that this is the asymptotic variance of $\sqrt{n}(t-\theta)$. Since $g_p'(\lambda)$ exists and is negative for all λ, it follows that $g_p^{-1}(\theta)$ is also differentiable. Hence, using 6a.2.1 of Rao [21], we have that

$$\sqrt{n}[g_p^{-1}(t) - g_p^{-1}(\theta)] = \sqrt{n}(\hat{\lambda}-\lambda)$$

is asymptotically normally distributed with zero mean and variance

$$V_p(\lambda) = [\frac{d}{d\theta} g_p^{-1}(\theta)]^2[\phi^2 + r^{-1}\gamma^2] .$$

Thus statements (b) and (c) are proved, and because $V_p(\lambda)$ does not depend on n, statement (a) also follows. □

There are several points which should be mentioned concerning the estimator $\hat{\lambda}$. First, it can easily be shown that $g_p^{-1}(\cdot)$ is a convex function, and thus by Jensen's inequality, it follows that $\hat{\lambda}$ is not an unbiased estimator for small samples, but rather $E\hat{\lambda} > \lambda$. In addition to being biased for small samples, the estimator also has a rather formidable exact distribution. However, the maximum likelihood estimator suffers from these same problems and $\hat{\lambda}$ is much easier to compute. $\hat{\lambda}$ also can be used with transformed data, and still will yield the valid estimate of λ. Second, there is a possibility of obtaining negative estimates of λ with $\hat{\lambda}$, while λ cannot be negative. One solution to this difficulty would be to replace any negative estimate of λ with $\hat{\lambda} = 0$. In any case, $\hat{\lambda} < 0$ only if $t > \frac{1}{2}$ and $\Pr(t > \frac{1}{2}) \to 0$ if $\lambda > 0$, due to the consistency of t. Thus this difficulty will be negligible for large n.

We would now like to examine the efficiency of $\hat{\lambda}$ with respect to λ^*. Certainly $\hat{\lambda}$ will be no more efficient than λ^*, since λ^* has asymptotically minimum variance. However, due to the other desirable properties of $\hat{\lambda}$, a statistically less efficient estimator could be acceptable. In the following section, we will investigate the asymptotic relative efficiency of λ and λ^*.

4.4 Asymptotic relative efficiency of λ and λ^*.

Since $\sqrt{n}\hat{\lambda}$ and $\sqrt{n}\lambda^*$ both have limiting normal distributions with zero mean and constant variance, we can use the ratio

$$e_p(\lambda) = \frac{V^*_p(\lambda)}{V_p(\lambda)} = \frac{\mathrm{Var}(\sqrt{n}\lambda^*)}{\mathrm{Var}(\sqrt{n}\hat{\lambda})}$$

as a measure of the efficiency of $\hat{\lambda}$ relative to λ^*. Expressions for the two variances involved in $e_p(\lambda)$ have been given in the previous two sections. Unfortunately, neither $V^*_p(\lambda)$ nor $V_p(\lambda)$ has a simple closed form. Consequently it is necessary to use numerical techniques to evaluate $e_p(\lambda)$. This section contains methods for the numerical evaluation of $V^*_p(\lambda)$ and $V_p(\lambda)$, and tables of these quantities for selected values of p and λ are included in the Appendix. We restrict our attention here to $\lambda \le 5$.

From Theorem 4.2.1, we have

$$V^*_p(\lambda) = \{-\frac{1}{4} + \frac{1}{8}\exp(-\tfrac{1}{2}\lambda)J^*_p(\lambda)\}^{-1}, \tag{4.4.1}$$

where

$$J^*_p(\lambda) = \int_0^\infty \left(\frac{u}{\lambda}\right)^{\frac{1}{2}q+1} \frac{I^2_{q+1}(\sqrt{\lambda u})}{I_q(\sqrt{\lambda u})} e^{-\frac{1}{2}u} du . \tag{4.4.2}$$

$I_0(x)$ can be approximated accurately by a polynomial (see, e.g., Abramowitz and Stegun [1], p. 378). Also $I_{\frac{1}{2}}(x) = [\sinh(x)]/x$. Using one of these expressions and a continued fraction based on the recurrence relation

$$I_{q+1}(x) = I_{q-1}(x) - 2(q/x)I_q(x) ,$$

we can calculate the value of $I_q(x)$ for any value of $q = \frac{1}{2}p-1$ when p is an integer greater than or equal to 2. Thus we can calculate the integrand in (4.4.2) for any values of q, λ and u we require. The value of $J^*_p(\lambda)$ can then be approximated using Gaussian quadrature with Laguerre polynomials (see, e.g., Abramowitz and Stegun [1], p. 890). Tables A.2(a) - (g) give values of $V^*_p(\lambda)$, computed as described above,

using 10-point quadrature, for $p = 2(1)8$ and $\lambda = 0(.1)2(.5)5$.

By Theorem 4.3.2, we have

$$V_p(\lambda) = [\frac{d}{d\theta} g_p^{-1}(\theta)]^2[\phi^2 + r^{-1}\gamma^2] . \tag{4.4.3}$$

Since we have already shown that $g_p'(\lambda) < 0$ for all λ, it follows that we can rewrite (4.4.3) as

$$V_p(\lambda) = \frac{\phi^2 + r^{-1}\gamma^2}{[g_p'(\lambda)]^2} . \tag{4.4.4}$$

By (4.3.3) and (4.3.4),

$$g_p'(\lambda) = (\tfrac{1}{2})^{p+1}[\Gamma(\tfrac{1}{2}p)]^{-1}e^{-\frac{1}{2}\lambda}\sum_{j=0}^{\infty}\frac{\Gamma(p+j)}{\Gamma(\frac{1}{2}p+j+1)j!}(\frac{\lambda}{4})^j .$$

In this form, $g_p'(\lambda)$ can be computed easily. Values of $g_p'(\lambda)$ are included in tables A.2(a) - (f). $\phi^2 + \theta^2$ and $\gamma^2 + \theta^2$ can be expressed as infinite series of the form $\sum_{j=0}^{\infty} q_j\lambda^j$, where the coefficients q_j satisfy certain recurrence relations. In this way, ϕ^2 and γ^2 can be computed numerically also. The exact forms of the series and the recurrence properties will be given in Theorem 4.4.1 for ϕ^2 and in Theorem 4.4.2 for γ^2. First, however, we need several lemmas which are based on the following two formulas for integration by parts:

$$\int \frac{x^n dx}{(\beta+x)^m} = \frac{n}{m-1}\int \frac{x^{n-1}dx}{(\beta+x)^{m-1}} - \frac{x^n}{(m-1)(\beta+x)^{m-1}} \tag{4.4.5}$$

$$\int \frac{x^n dx}{(\beta+x)^m} = \frac{m-n-2}{(m-1)\beta}\int \frac{x^n dx}{(\beta+x)^{m-1}} + \frac{x^{n+1}}{(m-1)\beta(\beta+x)^{m-1}} . \tag{4.4.6}$$

Both formulas are valid if $m > 1$.

Lemma 4.4.1. For $a, b \geq 0$ and $c \geq 1$,

$$\int_0^1\int_0^1 \frac{x^{a+1}y^b dx\, dy}{(1+x+y)^{a+b+c+1}} = \frac{a+1}{a+b+c}\int_0^1\int_0^1 \frac{x^a y^b dx\, dy}{(1+x+y)^{a+b+c}}$$

$$- \frac{\Gamma(b+1)\Gamma(a+c+1)}{\Gamma(a+b+c+1)} (\tfrac{1}{2})^{a+c+1} I_{1/3}(b+1,a+c-1) .$$

Proof:

$$\int_0^1\int_0^1 \frac{x^{a+1}y^b dx\, dy}{(1+x+y)^{a+b+c+1}} = \int_0^1 y^b J(y)\,dy , \qquad (4.4.7)$$

where

$$J(y) = \int_0^1 \frac{x^{a+1}dx}{(1+x+y)^{a+b+c+1}} .$$

Using (4.4.5) with $n = a+1$, $m = a+b+c+1$, $\beta = 1+y$, we obtain

$$J(y) = -\left[\frac{x^{a+1}}{(a+b+c)(1+x+y)^{a+b+c}}\right]_{x=0}^{1} + \frac{a+1}{a+b+c}\int_0^1 \frac{x^a dx}{(1+x+y)^{a+b+c}}$$

$$= -\frac{1}{(a+b+c)(2+y)^{a+b+c}} + \frac{a+1}{a+b+c}\int_0^1 \frac{x^a dx}{(1+x+y)^{a+b+c}} . \qquad (4.4.8)$$

Since

$$\int_0^1 \frac{y^b dy}{(2+y)^{a+b+c}} = (\tfrac{1}{2})^{a+c-1} \frac{\Gamma(b+1)\Gamma(a+c-1)}{\Gamma(a+b+c)} I_{1/3}(b+1,a+c-1) ,$$

the result follows immediately on substitution of (4.4.8) in (4.4.7). □

Lemma 4.4.2. For $a \geq 0$ and $c > \max(a,1)$,

$$\int_1^\infty\int_1^\infty \frac{x^a y^a dx\, dy}{(1+x+y)^{a+c-1}} = \frac{c-a-2}{a+c}\int_1^\infty\int_1^\infty \frac{x^a y^a dx\, dy}{(1+x+y)^{a+c}}$$

$$- (\tfrac{1}{2})^{c-2} \frac{\Gamma(a+1)\Gamma(c-1)}{\Gamma(a+c+1)} I_{2/3}(c-1,a+1) .$$

Proof: Using (4.4.6) with $n = a$, $m = a+c+1$, $\beta = 1+y$, we obtain the following upon multiplying both sides by $(1+y)y^a$ and integrating with respect to y:

$$\int_1^\infty \int_1^\infty \frac{x^a y^a dx\, dy}{(1+x+y)^{a+c+1}} + \int_1^\infty \int_1^\infty \frac{x^a y^{a+1} dx\, dy}{(1+x+y)^{a+c+1}}$$

$$= \frac{c-1}{a+c} \int_1^\infty \int_1^\infty \frac{x^a y^a dx\, dy}{(1+x+y)^{a+c}} - \frac{1}{a+c} \int_1^\infty \frac{y^a dy}{(2+y)^{a+c}} . \qquad (4.4.9)$$

Substituting y for x in (4.4.5), then letting $n = a+1$, $m = a+c$, $\beta = 1+x$, we obtain the following upon multiplying both sides by x^a and integrating with respect to x:

$$\int_1^\infty \int_1^\infty \frac{x^a y^{a+1} dx\, dy}{(1+x+y)^{a+c+1}} = \frac{a+1}{a+c} \int_1^\infty \int_1^\infty \frac{x^a y^a dx\, dy}{(1+x+y)^{a+c}}$$

$$+ \frac{1}{a+c} \int_1^\infty \frac{x^a\, dx}{(2+x)^{a+c}} . \qquad (4.4.10)$$

Substituting (4.4.10) in (4.4.9), we obtain

$$\int_1^\infty \int_1^\infty \frac{x^a y^a dx\, dy}{(1+x+y)^{a+c+1}} = \left(\frac{c-1}{a+c} - \frac{a+1}{a+c}\right) \int_1^\infty \int_1^\infty \frac{x^a y^a dx\, dy}{(1+x+y)^{a+c}}$$

$$- \frac{2}{a+c} \int_1^\infty \frac{y^a dy}{(2+y)^{a+c}} . \qquad (4.4.11)$$

Since

$$\int_1^\infty \frac{y^a dy}{(2+y)^{a+c}} = (\tfrac{1}{2})^{c-1} \frac{\Gamma(a+1)\Gamma(c-1)}{\Gamma(a+c)} I_{2/3}(c-1,a+1),$$

the result follows upon substitution in (4.4.11) and simplification.

$\square$

We will also need the form of the density function of the multivariate inverted beta distribution. It is given in Johnson and Kotz [12], p. 238, and is included here for reference:

If $X_0, X_1, \ldots, X_n$ are independent random variables with X_j distributed as $\chi^2_{\nu_j}$ $(j = 0,1,\ldots,n)$, then $Y_j = X_j/X_0$ $(j = 1,\ldots,n)$ have a multivariate inverted beta distribution with density function

$$p_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = \frac{\Gamma(\frac{1}{2}\nu)}{\prod_{j=0}^{n}\Gamma(\frac{1}{2}\nu_j)} \frac{\prod_{j=1}^{n} y_j^{\frac{1}{2}\nu_j - 1}}{\left(1+\sum_{j=1}^{n} y_j\right)^{\frac{1}{2}\nu}},$$

$$(y_j \geq 0, j=1,\ldots,n) \qquad (4.4.12)$$

where

$$\nu = \textstyle\sum_{j=0}^{n} \nu_j \; .$$

We now proceed with the series representations of $\phi^2 + \theta^2$ and $\gamma^2 + \theta^2$.

Theorem 4.4.1. Using the notation of Theorem 4.3.1, $\phi^2 + \theta^2$ can be expressed as

$$\phi^2 + \theta^2 = e^{-\frac{1}{2}\lambda} \sum_{j=0}^{\infty} q_j (\tfrac{1}{2}\lambda)^j / j! \;, \qquad (4.4.13)$$

where $q_0 = 1/3$ and for $j = 0,1,\ldots$, the following recurrence relation is satisfied:

$$q_{j+1} = q_j - \frac{\Gamma(p+j)}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}p+j+1)} (\tfrac{1}{2})^{p+j-1} I_{2/3}(p+j, \tfrac{1}{2}p). \qquad (4.4.14)$$

Further, if R_n is the remainder after n+1 terms in (4.4.13), then

$$|R_n| < (\tfrac{1}{2}\lambda)^{n+1} q_{n+1}/(n+1)! \;. \qquad (4.4.15)$$

<u>Proof</u>: We have $\phi^2 + \theta^2 = \Pr(X_1<Y_1, X_1<Y_2)$, where X_1 is distributed as $\chi'^2_p(\lambda)$ and Y_1, Y_2 are distributed as χ^2_p and all three are independent. Since the density of X_1 has the representation (4.1.1), it is clear that

we can rewrite $\phi^2 + \theta^2$ as in (4.4.13) where $q_j = \Pr(U_j<Y_1,\ U_j<Y_2)$, Y_1 and Y_2 being as before and U_j being distributed independently as χ^2_{p+2j}. Thus $q_j = \Pr(Z_j^{(1)}>1,\ Z_j^{(2)}>1)$, where $Z_j^{(1)} = Y_1/U_j$ and $Z_j^{(2)} = Y_2/U_j$. But $Z_j^{(1)}$ and $Z_j^{(2)}$ have a bivariate inverted beta distribution. Thus from (4.4.12), with $\nu_0 = p+2j+2$, $\nu_1 = \nu_2 = p$, we obtain

$$q_{j+1} = \frac{\Gamma((3/2)p+j+1)}{\Gamma(\tfrac{1}{2}p+j+1)[\Gamma(\tfrac{1}{2}p)]^2}\int_1^\infty\int_1^\infty \frac{x^{\frac{1}{2}p-1}y^{\frac{1}{2}p-1}}{(1+x+y)^{(3/2)p+j+1}}\,dx\,dy\,.$$

Using Lemma 4.4.2 with $a = \tfrac{1}{2}p-1$, $c = p+j+1$, we obtain

$$q_{j+1} = \frac{\Gamma((3/2)p+j+1)}{\Gamma(\tfrac{1}{2}p+j+1)[\Gamma(\tfrac{1}{2}p)]^2}\left\{\frac{(\tfrac{1}{2}p+j)}{((3/2)p+j)}\int_1^\infty\int_1^\infty \frac{x^{\frac{1}{2}p-1}y^{\frac{1}{2}p-1}\,dx\,dy}{(1+x+y)^{(3/2)p+j}}\right.$$

$$\left. - \frac{\Gamma(\tfrac{1}{2}p)\Gamma(p+j)}{\Gamma((3/2)p+j+1)}(\tfrac{1}{2})^{p+j-1} I_{2/3}(p+j,\tfrac{1}{2}p)\right\}.$$

(4.4.14) follows immediately on simplification.

When $j = 0$, we have

$$q_0 = \Pr(U_0<Y_1,\ U_0<Y_2)$$

where U_0, Y_1, Y_2 are i.i.d. Clearly

$$\Pr(U_0<Y_1,\ U_0<Y_2) = \Pr(Y_1<U_0,\ Y_1<Y_2) = \Pr(Y_2<U_0,\ Y_2<Y_1)\,.$$

Hence, $q_0 = 1/3$. Since q_j decreases as j increases, clearly

$$|R_n| = e^{-\frac{1}{2}\lambda}\left|\sum_{j=n+1}^{\infty} q_j \frac{(\tfrac{1}{2}\lambda)^j}{j!}\right|$$

$$< e^{-\frac{1}{2}\lambda} q_{n+1}\left|\sum_{j=n+1}^{\infty}\frac{(\tfrac{1}{2}\lambda)^j}{j!}\right|\,.$$

But $\sum_{j=n+1}^{\infty}\frac{(\tfrac{1}{2}\lambda)^j}{j!}$ is the remainder after n+1 terms in the Taylor series

expansion for $e^{\frac{1}{2}\lambda}$. Thus for $\lambda \geq 0$,

$$\left|\sum_{j=n+1}^{\infty} \frac{(\frac{1}{2}\lambda)^j}{j!}\right| < (\tfrac{1}{2}\lambda)^{n+1} e^{-\frac{1}{2}\lambda}/(n+1)!$$

and (4.4.15) follows immediately. □

<u>*Theorem 4.4.2*</u>. Using the notation of Theorem 4.3.1, $\gamma^2 + \theta^2$ can expressed as

$$\gamma^2 + \theta^2 = e^{-\lambda} \sum_{j=0}^{\infty} q_j (\tfrac{1}{2}\lambda)^j , \tag{4.4.16}$$

where

$$q_j = \sum_{k=0}^{j} \frac{p_{j-k,k}}{(j-k)!k!} ,$$

and the quantities $p_{j,k}$ satisfy the following recurrence relations:

(i) $p_{0,0} = 1/3$

$$\text{(ii)} \quad p_{j+1,k} = p_{j,k} - \frac{\Gamma(p+j)}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}p+j+1)} (\tfrac{1}{2})^{p+j} I_{1/3}(\tfrac{1}{2}p+k,p+j) \tag{4.4.17}$$

$$\text{(iii)} \quad p_{j+1,j+1} = p_{j,j} - \frac{\Gamma(p+j)}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}p+j+1)} (\tfrac{1}{2})^{p+j}$$

$$\times \left\{I_{1/3}(\tfrac{1}{2}p+j,p+j) + I_{1/3}(\tfrac{1}{2}p+j+1,p+j)\right\} , \tag{4.4.18}$$

(ii) holding for $j,k = 0,1,\ldots$ and (iii) holding for $j = 0,1,\ldots$. Further, if R_n is the remainder after n+1 terms in (4.4.16), then

$$|R_n| < p^*_{n+1} \lambda^{n+1}/(n+1)! , \tag{4.4.19}$$

where $p^*_n = \max\limits_{1 \leq k \leq n} p_{n-k,k}$.

<u>Proof</u>: $\gamma^2 + \theta^2 = \Pr(X_1 < Y_1,\ X_2 < Y_1)$ where X_1 and X_2 are distributed as $\chi'^2_p(\lambda)$ and Y_1 is distributed as χ^2_p, and all three are independent. Since the densities of both X_1 and X_2 have the representation (4.1.1), we can

rewrite $\gamma^2 + \theta^2$ as

$$\gamma^2 + \theta^2 = e^{-\lambda} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{(\frac{1}{2}\lambda)^j (\frac{1}{2}\lambda)^k}{j!k!} \Pr(U_j^{(1)} < Y_1,\ U_k^{(2)} < Y_1) \ ,$$

where $U_j^{(1)}$, $U_k^{(2)}$ and Y_1 are independently distributed as χ^2_{p+2j}, χ^2_{p+2k} and χ^2_p respectively. Letting $p_{j,k} = \Pr(U_j^{(1)} < Y_1,\ U_k^{(2)} < Y_1)$ and rearranging terms in the sum, we immediately obtain (4.4.16). Since $U_0^{(1)}$, $U_0^{(2)}$ and Y_1 are i.i.d., $p_{0,0} = 1/3$ as was q_0 in Theorem 4.4.1. We also have $p_{j,k} = \Pr(Z_j^{(1)} < 1,\ Z_k^{(2)} < 1)$ where $Z_j^{(1)} = U_j^{(1)}/Y_1$, $Z_j^{(2)} = U_j^{(2)}/Y_1$, and thus $Z_j^{(1)}$ and $Z_j^{(2)}$ have a bivariate inverted beta distribution. Thus from (4.4.12) with $\nu_0 = p$, $\nu_1 = p+2j+2$, $\nu_2 = p+2k$, we obtain

$$p_{j+1,k} = \frac{\Gamma((3/2)p+j+k+1)}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}p+j+1)\Gamma(\frac{1}{2}p+k)} \int_0^1 \int_0^1 \frac{x^{\frac{1}{2}p+j} y^{\frac{1}{2}p+k-1}}{(1+x+y)^{(3/2)p+j+k+1}} dx\, dy \ .$$

Using Lemma 4.4.1 with $a = \frac{1}{2}p+j-1$, $b = \frac{1}{2}p+k=1$, $c = \frac{1}{2}p+2$, we obtain

$$p_{j+1,k} = \frac{\Gamma((3/2)p+j+k+1)}{(\frac{1}{2}p)\Gamma(\frac{1}{2}p+j+1)\Gamma(\frac{1}{2}p+k)} \left\{ \frac{(\frac{1}{2}p+j)}{((3/2)p+j+k)} \int_0^1 \int_0^1 \frac{x^{\frac{1}{2}p+j-1} y^{\frac{1}{2}p+k-1}}{(1+x+y)^{(3/2)p+j+k+1}} dx\, dy \right.$$

$$\left. - \frac{\Gamma(\frac{1}{2}p+k)\Gamma(p+j)}{\Gamma((3/2)p+j+k+1)} (\tfrac{1}{2})^{p+j} I_{1/3}(\tfrac{1}{2}p+k, p+j) \right\} .$$

(4.4.17) follows immediately on simplification. (4.4.18) follows similarly using Lemma 4.4.1 twice, first as it stands with $a = \frac{1}{2}p+j-1$, $b = \frac{1}{2}p+j$, $c = \frac{1}{2}p+2$, and then with x and y interchanged and $a = \frac{1}{2}p+j-1$, $b = \frac{1}{2}p+j-1$, $c = \frac{1}{2}p+2$.

For any k, $p_{j,k}$ decreases as j increases. Thus for $j \geq n+1$ and $k \leq j$, $p_{j-k,k} \leq p^*_{n-1}$, with strict equality at least when $j > n+1$. Thus

$$|R_n| = e^{-\lambda} \left| \sum_{j=n+1}^{\infty} \sum_{k=0}^{j} \frac{(\frac{1}{2}\lambda)^j}{(j-k)!k!} p_{j-k,k} \right| <$$

$$< e^{-\lambda} p^*_{n+1} \left| \sum_{j=n+1}^{\infty} (\tfrac{1}{2}\lambda)^j \sum_{k=0}^{j} \frac{1}{(j-k)!k!} \right| .$$

But

$$\sum_{k=0}^{j} \frac{1}{(j-k)!k!} = \frac{1}{j!} \sum_{k=0}^{j} \binom{j}{k} = \frac{2j}{j!} .$$

Thus

$$|R_n| < e^{-\lambda} p^*_{n+1} \left| \sum_{j=n+1}^{\infty} \frac{\lambda^j}{j!} \right| .$$

Since $\sum_{j=n+1}^{\infty} (\lambda^j)/j!$ is the remainder after n+1 terms in the Taylor series expansion for e^{λ}, it follows that for $\lambda \geq 0$,

$$\left| \sum_{j=n+1}^{\infty} \frac{\lambda^j}{j!} \right| < \frac{\lambda^{n+1}}{(n+1)!} e^{-\lambda} ,$$

and (4.4.19) follows upon substitution. □

Using the recurrence properties outlined in the above theorems, computation of ϕ^2 and γ^2 is straightforward. These quantities are also included in Tables A.2(a) - (g).

The quantities ϕ^2, γ^2, and $g'_p(\lambda)$ are sufficient for calculating $V_p(\lambda)$ and hence $e_p(\lambda)$ for any value of r, where $r = \lim_{n\to\infty}(m/n)$. Tables A.3(a) - (g) give $e_p(\lambda)$ for $\lambda = 0(.1)2(.5)5$ and $r = 1,2,4,8$.

Several features of the tables deserve specific mention. We notice, for example, that for fixed p and λ, there is a rapid increase in efficiency as r increases from 1 to 4. As r increases further, there is additional increase in efficiency, but it is more gradual. We also note that for fixed p and r, $e_p(\lambda)$ gradually rises to a maximum as λ increases; then $e_p(\lambda)$ decreases, slowly at first, more rapidly for larger λ. However, in each case, the efficiency is relatively stable for values of λ up to 2 or 3.

As mentioned in section 4.1, the need for estimating the noncentral-

ity parameter often arises in problems concerning distances. If we consider the distances between points in the unit sphere, then the values of λ which are of interest are in the range 0 to 2, where the efficiency is relatively constant.

A third feature of interest in the tables is the value of λ for which maximum efficiency is attained. For values of p considered here, that value is about 2 when r = 1. As r increases, the value where maximum efficiency is attained decreases, rapidly at first, then more slowly for larger r. The maximal efficiency for each combination of p and r is indicated by an asterisk in Tables A.3(a) - (g) to facilitate these comparisons.

In summary, we find that the randomized estimator $\hat{\lambda}$ is less efficient than the maximum likelihood estimator λ^* when the data are appropriate for the use of the λ^*. However, for moderate values of r, the loss of efficiency is not too severe, and $\hat{\lambda}$ is easier to compute than λ^*. $\hat{\lambda}$ also can be utilized when the data have been subjected to a monotonic distortion if control observations subject to the same distortion are available. Thus it is believed that the general procedure and specific estimator which results should be of interest and value.

4.5 Extensions and generalizations.

The most obvious generalization of the results given here is the removal of the assumption that the sample observations have come from a noncentral chi-squared distribution. In fact, the general method used here would be applicable anytime that Pr(X<Y) can be expressed as an invertible function of an unknown parameter. Perhaps the method could be applied in other situations where standard procedures lead to intractable estimators.

Another desirable generalization would be the removal of the assumption of independence among the observations. If this could be done, the results of this chapter might be applied directly to problems arising under models such as those considered in Chapters two and three. The requirement of independent observations hampers those applications of the theory developed here. Work has been done on limit theory under various dependence assumptions, and that may be applicable to statistics of the form considered here. However, the author is unaware of such results when the two samples cannot even be considered independent of each other. Further investigation along these lines would be of interest and is being considered.

BIBLIOGRAPHY

1. Abramowitz, M., and Stegun, I.A. (1970), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, Applied Mathematics Series, no. 55.

2. Anderson, T.W. (1958), *An Introduction to Multivariate Analysis*, New York, John Wiley and Sons, Inc.

3. Cormack, R.M. (1971), "A review of classification," *Journal of the Royal Statistical Society*, Series A, 134, 321-353.

4. Das Gupta, S. (1964), "Non-parametric classification rules," *Sankhyā*, Series A, 26, 25-30.

5. David, H.A. (1963), *The Method of Paired Comparisons*, London, Chas. Griffin and Co., Ltd.

6. Fraser, D.A.S. (1957), *Nonparametric Methods in Statistics*, New York, John Wiley and Sons, Inc.

7. Gradshteyn, I.S., and Ryzhik, I.M. (1965), *Table of Integrals, Series and Products*, New York, Academic Press.

8. Gurland, J. (1953), "Distribution of quadratic forms and ratios of quadratic forms," *Annals of Mathematical Statistics*, 24, 416-427.

9. Gurland, J. (1956), "Quadratic forms in normally distributed random variables," *Sankhyā*, 17, 37-50.

10. Hartigan, J.A. (1975), *Clustering Algorithms*, New York, John Wiley and Sons, Inc.

11. Johnson, N.L., and Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions - 2*, Boston, Houghton Mifflin.

12. Johnson, N.L., and Kotz, S. (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, New York, John Wiley and Sons.

13. Kendall, M.G., and Stuart, A. (1963), *The Advanced Theory of Statistics*, Vol. 1, London, Charles Griffin and Co., Ltd.

14. Kendall, M.G., and Stuart, A. (1967), *The Advanced Theory of Statistics*, Vol. 2, New York, Hafner Publishing Co.

15. Kendall, M.G., and Stuart, A. (1967), *The Advanced Theory of Statistics*, Vol. 3, New York, Hafner Publishing Co.

16. Kruskal, J.B. (1964), "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, 29, 1-27.

17. Kruskal, J.B. (1964), "Nonmetric multidimensional scaling: A numerical method," *Psychometrika,* 29, 115-129.

18. Meyer, P.L. (1967), "The maximum likelihood estimate of the noncentrality parameter of a noncentral chi-squared variate," *Journal of the American Statistical Association,* 62, 1258-1264.

19. Pandey, J.N., and Rahman, M. (1971), "The maximum likelihood estimate of the noncentrality parameter of a noncentral F variate," *SIAM Journal of Mathematical Analysis,* 2, 269-276.

20. Press, S.J. (1966), "Linear combinations of non-central chi-square variates," *Annals of Mathematical Statistics,* 37, 480-487.

21. Rao, C.R. (1965), *Linear Statistical Inference and its Applications,* New York, John Wiley and Sons, Inc.

22. Robbins, H. (1948), "The distribution of a definite quadratic form," *Annals of Mathematical Statistics,* 19, 266-270.

23. Shepard, R.N. (1962), "The analysis of proximities: Multidimensional scaling with an unknown distance function, I," *Psychometrika,* 27, 125-140.

24. Shepard, R.N. (1962), "The analysis of proximities: Multidimensional scaling with an unknown distance function, II," *Psychometrika,* 27, 219-246.

25. Shepard, R.N. (1972), "A taxonomy of some principal types of data and of multidimensional methods for their analysis," *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences,* R.N. Shepard, A.K. Romney, and S.B. Nerlove, eds., Vol. 1, 21-47, New York, Seminar Press.

26. Van Dantzig, D. (1951), "On the consistency and the power of Wilcoxon's two sample test," *Nederlandse Akademie van Wetenschappen,* Proceedings, Series A, 54, 1-8.

APPENDIX

TABLE A.1: $g_p^{-1}(\theta)$

θ	p = 2	3	4	5	6	7	8
.50	0.0	0.0	0.0	0.0	0.0	0.0	0.0
.49	.081	.095+	.108	.119	.129	.139	.147
.48	.163	.192	.217	.240	.260	.279	.297
.47	.248	.291	.329	.363	.394	.422	.449
.46	.334	.392	.443	.488	.529	.568	.603
.45	.421	.495+	.559	.616	.668	.715+	.760
.44	.511	.600	.677	.746	.808	.866	.920
.43	.603	.708	.798	.878	.952	1.019	1.082
.42	.697	.818	.921	1.014	1.098	1.175+	1.248
.41	.794	.930	1.047	1.152	1.247	1.335-	1.416
.40	.893	1.045+	1.176	1.293	1.399	1.497	1.588
.39	.994	1.163	1.308	1.437	1.554	1.663	1.763
.38	1.098	1.284	1.443	1.585-	1.713	1.832	1.942
.37	1.204	1.408	1.581	1.736	1.876	2.005-	2.125-
.36	1.314	1.535-	1.723	1.890	2.042	2.181	2.312
.35	1.427	1.665-	1.869	2.049	2.212	2.362	2.502
.34	1.543	1.799	2.018	2.211	2.386	2.548	2.698
.33	1.662	1.937	2.171	2.378	2.565+	2.738	2.898
.32	1.785+	2.079	2.329	2.549	2.749	2.933	3.104
.31	1.912	2.225+	2.491	2.726	2.938	3.133	3.314
.30	2.043	2.376	2.658	2.907	3.132	3.339	3.531
.29	2.179	2.532	2.830	3.094	3.332	3.550+	3.754
.28	2.319	2.693	3.008	3.287	3.538	3.768	3.983
.27	2.465-	2.859	3.192	3.486	3.750+	3.993	4.219
.26	2.616	3.032	3.383	3.692	3.971	4.226	4.463
.25	2.773	3.211	3.580	3.905-	4.197	4.466	4.715+
.24	2.936	3.397	3.785+	4.126	4.433	4.715-	4.976
.23	3.106	3.591	3.998	4.356	4.678	4.973	5.247
.22	3.284	3.793	4.220	4.595-	4.932	5.241	5.528
.21	3.470	4.004	4.452	4.844	5.197	5.520	5.821

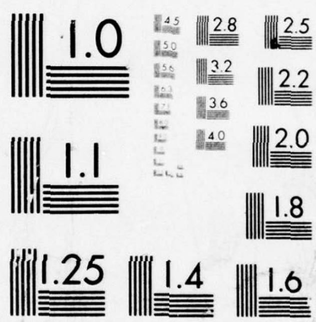
Table A.1 (continued)

θ	p = 2	3	4	5	6	7	8
.20	3.665+	4.225+	4.694	5.105-	5.474	5.812	6.126
.19	3.870	4.457	4.948	5.377	5.763	6.117	6.444
.18	4.087	4.702	5.215+	5.664	6.067	6.436	6.778
.17	4.315+	4.959	5.496	5.965+	6.387	6.772	7.129
.16	4.558	5.232	5.794	6.284	6.724	7.126	7.500-
.15	4.816	5.522	6.110	6.622	7.081	7.501	7.891
.14	5.092	5.832	6.446	6.981	7.461	7.900	8.307
.13	5.388	6.163	6.806	7.366	7.867	8.326	8.750+
.12	5.708	6.521	7.194	7.779	8.303	8.783	9.226
.11	6.057	6.909	7.613	8.226	8.775-	9.276	9.740
.10	6.438	7.333	8.071	8.713	9.288	9.813	10.299

AD-A051 753 NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS F/G 12/1
SOME STATISTICAL PROCEDURES BASED ON DISTANCES.(U)
NOV 76 J J WALKER DAAG29-74-C-0030
UNCLASSIFIED MMS-1096 ARO-11959.19-M NL

2 OF 2
AD
A051753

END
DATE
FILMED
4-78
DDC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

TABLE A.2(a): Factors for Efficiency Computations (p = 2)

λ	$g_p(\lambda)$	$g'_p(\lambda)$	ϕ^2	γ^2	$V^*_p(\lambda)$
0.0	.500	-.1250	.0833	.0833	4.000
0.1	.488	-.1219	.0846	.0819	4.392
0.2	.476	-.1189	.0856	.0806	4.771
0.3	.464	-.1160	.0864	.0792	5.143
0.4	.452	-.1131	.0870	.0778	5.510
0.5	.441	-.1103	.0875	.0764	5.874
0.6	.430	-.1076	.0877	.0750	6.235
0.7	.420	-.1049	.0878	.0736	6.595
0.8	.409	-.1023	.0877	.0723	6.954
0.9	.399	-.0998	.0875	.0709	7.313
1.0	.389	-.0974	.0872	.0695	7.671
1.1	.380	-.0950	.0868	.0682	8.030
1.2	.370	-.0926	.0862	.0668	8.389
1.3	.361	-.0903	.0856	.0655	8.748
1.4	.352	-.0881	.0848	.0641	9.108
1.5	.344	-.0859	.0841	.0628	9.469
1.6	.335	-.0838	.0832	.0615	9.831
1.7	.327	-.0817	.0823	.0602	10.193
1.8	.319	-.0797	.0813	.0589	10.557
1.9	.311	-.0777	.0802	.0576	10.921
2.0	.303	-.0758	.0792	.0564	11.286
2.5	.268	-.0669	.0732	.0503	13.125
3.0	.236	-.0590	.0668	.0445	14.986
3.5	.208	-.0521	.0603	.0393	16.868
4.0	.184	-.0460	.0539	.0344	18.766
4.5	.162	-.0406	.0479	.0300	20.680
5.0	.143	-.0358	.0422	.0261	22.606

TABLE A.2(b): Factors for Efficiency Computations (p = 3)

λ	$g_p(\lambda)$	$g'_p(\lambda)$	ϕ^2	γ^2	$V^*_p(\lambda)$
0.0	.500	-.1061	.0833	.0833	6.000
0.1	.489	-.1040	.0842	.0824	6.395
0.2	.479	-.1019	.0849	.0814	6.783
0.3	.469	-.0999	.0855	.0804	7.165
0.4	.459	-.0979	.0860	.0793	7.544
0.5	.450	-.0960	.0863	.0783	7.920
0.6	.440	-.0940	.0865	.0772	8.293
0.7	.431	-.0921	.0865	.0762	8.666
0.8	.422	-.0903	.0865	.0751	9.036
0.9	.413	-.0885	.0864	.0740	9.407
1.0	.404	-.0870	.0862	.0729	9.776
1.1	.395	-.0849	.0859	.0718	10.146
1.2	.387	-.0832	.0855	.0707	10.515
1.3	.379	-.0815	.0850	.0696	10.884
1.4	.371	-.0799	.0845	.0685	11.253
1.5	.363	-.0783	.0839	.0673	11.623
1.6	.355	-.0767	.0832	.0662	11.993
1.7	.347	-.0751	.0825	.0651	12.363
1.8	.340	-.0736	.0818	.0640	12.734
1.9	.333	-.0721	.0810	.0628	13.014
2.0	.326	-.0706	.0801	.0617	13.476
2.5	.292	-.0636	.0754	.0562	15.341
3.0	.262	-.0573	.0701	.0508	17.219
3.5	.235	-.0515	.0645	.0457	19.111
4.0	.210	-.0463	.0589	.0409	21.015
4.5	.188	-.0416	.0533	.0363	22.930
5.0	.168	-.0374	.0480	.0321	24.855

TABLE A.2(c): Factors for Efficiency Computations (p = 4)

λ	$g_p(\lambda)$	$g'_p(\lambda)$	ϕ^2	γ^2	$V^*_p(\lambda)$
0.0	.500	-.0938	.0833	.0833	8.000
0.1	.491	-.0922	.0840	.0826	8.397
0.2	.482	-.0907	.0846	.0818	8.789
0.3	.473	-.0892	.0850	.0810	9.177
0.4	.464	-.0877	.0854	.0802	9.562
0.5	.455	-.0862	.0856	.0794	9.945
0.6	.446	-.0847	.0858	.0785	10.326
0.7	.438	-.0833	.0859	.0776	10.705
0.8	.430	-.0819	.0858	.0767	11.083
0.9	.422	-.0805	.0858	.0758	11.461
1.0	.414	-.0791	.0856	.0749	11.838
1.1	.406	-.0777	.0854	.0740	12.214
1.2	.398	-.0764	.0851	.0730	12.590
1.3	.391	-.0751	.0847	.0721	12.965
1.4	.383	-.0738	.0843	.0711	13.341
1.5	.376	-.0725	.0838	.0701	13.716
1.6	.369	-.0712	.0833	.0691	14.092
1.7	.362	-.0700	.0827	.0681	14.468
1.8	.355	-.0687	.0821	.0672	14.843
1.9	.348	-.0675	.0815	.0662	15.219
2.0	.341	-.0663	.0808	.0652	15.596
2.5	.309	-.0606	.0768	.0601	17.481
3.0	.280	-.0554	.0723	.0551	19.374
3.5	.254	-.0505	.0674	.0503	21.276
4.0	.230	-.0460	.0624	.0455	23.187
4.5	.208	-.0418	.0573	.0411	25.106
5.0	.188	-.0381	.0523	.0368	27.034

TABLE A.2(d): Factors for Efficiency Computations (p = 5)

λ	$g_p(\lambda)$	$g'_p(\lambda)$	ϕ^2	γ^2	$V^*_p(\lambda)$
0.0	.500	-.0849	.0833	.0833	10.000
0.1	.492	-.0837	.0839	.0827	10.398
0.2	.483	-.0825	.0843	.0821	10.792
0.3	.475	-.0813	.0847	.0814	11.184
0.4	.467	-.0801	.0850	.0808	11.573
0.5	.459	-.0789	.0852	.0800	11.960
0.6	.451	-.0778	.0854	.0793	12.345
0.7	.443	-.0766	.0854	.0786	12.729
0.8	.436	-.0755	.0854	.0778	13.112
0.9	.428	-.0744	.0854	.0770	13.495
1.0	.421	-.0733	.0852	.0762	13.876
1.1	.414	-.0722	.0851	.0754	14.257
1.2	.407	-.0711	.0848	.0746	14.638
1.3	.400	-.0700	.0845	.0738	15.018
1.4	.393	-.0690	.0842	.0729	15.398
1.5	.386	-.0679	.0838	.0720	15.778
1.6	.379	-.0669	.0834	.0712	16.158
1.7	.372	-.0658	.0829	.0703	16.537
1.8	.366	-.0648	.0824	.0694	16.917
1.9	.359	-.0638	.0818	.0685	17.297
2.0	.353	-.0628	.0813	.0676	17.677
2.5	.323	-.0580	.0779	.0630	19.578
3.0	.295	-.0535	.0739	.0583	21.484
3.5	.269	-.0493	.0695	.0537	23.396
4.0	.246	-.0453	.0650	.0492	25.314
4.5	.224	-.0416	.0603	.0448	27.238
5.0	.204	-.0382	.0556	.0406	29.169

TABLE A.2(e): Factors for Efficiency Computations (p = 6)

λ	$g_p(\lambda)$	$g'_p(\lambda)$	ϕ^2	γ^2	$V^*_p(\lambda)$
0.0	.500	-.0781	.0833	.0833	12.000
0.1	.492	-.0771	.0838	.0828	12.399
0.2	.485	-.0762	.0842	.0823	12.794
0.3	.477	-.0752	.0845	.0817	13.188
0.4	.470	-.0743	.0847	.0811	13.579
0.5	.462	-.0733	.0849	.0805	13.969
0.6	.455	-.0724	.0851	.0799	14.358
0.7	.448	-.0714	.0851	.0793	14.745
0.8	.441	-.0705	.0851	.0786	15.132
0.9	.434	-.0696	.0851	.0779	15.518
1.0	.427	-.0686	.0850	.0772	15.902
1.1	.420	-.0677	.0848	.0765	16.287
1.2	.413	-.0668	.0846	.0757	16.671
1.3	.406	-.0659	.0844	.0750	17.055
1.4	.400	-.0650	.0841	.0742	17.438
1.5	.393	-.0641	.0838	.0734	17.821
1.6	.387	-.0633	.0834	.0727	18.204
1.7	.381	-.0624	.0830	.0719	18.587
1.8	.375	-.0615	.0826	.0711	18.970
1.9	.369	-.0607	.0821	.0702	19.352
2.0	.362	-.0598	.0816	.0694	19.735
2.5	.334	-.0557	.0786	.0652	21.649
3.0	.307	-.0518	.0751	.0608	23.565
3.5	.282	-.0481	.0712	.0565	25.486
4.0	.259	-.0445	.0670	.0522	27.411
4.5	.237	-.0412	.0627	.0479	29.341
5.0	.217	-.0381	.0583	.0438	31.275

TABLE A.2(f): Factors for Efficiency Computations (p = 7)

λ	$g_p(\lambda)$	$g'_p(\lambda)$	ϕ^2	γ^2	$V^*_p(\lambda)$
0.0	.500	-.0728	.0833	.0833	14.000
0.1	.493	-.0719	.0837	.0829	14.399
0.2	.486	-.0711	.0841	.0824	14.796
0.3	.479	-.0703	.0843	.0819	15.191
0.4	.472	-.0695	.0846	.0814	15.584
0.5	.465	-.0687	.0847	.0809	15.976
0.6	.458	-.0679	.0848	.0803	16.367
0.7	.451	-.0672	.0849	.0798	16.756
0.8	.444	-.0664	.0849	.0792	17.145
0.9	.438	-.0656	.0849	.0785	17.534
1.0	.431	-.0648	.0848	.0779	17.921
1.1	.425	-.0640	.0847	.0773	18.308
1.2	.418	-.0633	.0845	.0766	18.695
1.3	.412	-.0625	.0843	.0759	19.081
1.4	.406	-.0617	.0841	.0752	19.467
1.5	.400	-.0610	.0838	.0745	19.853
1.6	.394	-.0602	.0835	.0738	20.238
1.7	.388	-.0595	.0831	.0731	20.623
1.8	.382	-.0587	.0827	.0724	21.009
1.9	.376	-.0580	.0823	.0716	21.393
2.0	.370	-.0572	.0818	.0708	21.779
2.5	.343	-.0536	.0729	.0669	23.703
3.0	.317	-.0502	.0760	.0629	25.628
3.5	.292	-.0469	.0725	.0587	27.557
4.0	.270	-.0437	.0686	.0546	29.487
4.5	.249	-.0407	.0646	.0505	31.422
5.0	.229	-.0379	.0604	.0465	33.361

TABLE A.2(g): Factors for Efficiency Computations (p = 8)

λ	$g_p(\lambda)$	$g'_p(\lambda)$	ϕ^2	γ^2	$V^*_p(\lambda)$
0.0	.500	-.0684	.0833	.0833	16.000
0.1	.493	-.0677	.0837	.0829	16.400
0.2	.486	-.0670	.0840	.0825	16.797
0.3	.480	-.0663	.0842	.0821	17.193
0.4	.473	-.0656	.0844	.0816	17.587
0.5	.467	-.0649	.0846	.0812	17.981
0.6	.460	-.0643	.0847	.0807	18.373
0.7	.454	-.0636	.0847	.0802	18.765
0.8	.447	-.0629	.0847	.0796	19.155
0.9	.441	-.0622	.0847	.0791	19.546
1.0	.435	-.0616	.0846	.0785	19.935
1.1	.429	-.0609	.0845	.0779	20.324
1.2	.423	-.0602	.0844	.0773	20.713
1.3	.417	-.0596	.0842	.0767	20.101
1.4	.411	-.0589	.0840	.0761	21.489
1.5	.405	-.0582	.0838	.0754	21.877
1.6	.399	-.0576	.0835	.0747	22.264
1.7	.394	-.0569	.0832	.0741	22.651
1.8	.388	-.0563	.0828	.0734	23.038
1.9	.382	-.0556	.0824	.0727	23.425
2.0	.377	-.0550	.0820	.0720	23.812
2.5	.350	-.0518	.0797	.0683	25.745
3.0	.325	-.0487	.0768	.0645	27.678
3.5	.301	-.0457	.0735	.0606	29.613
4.0	.279	-.0429	.0699	.0567	31.550
4.5	.259	-.0401	.0661	.0528	33.489
5.0	.239	-.0375	.0623	.0489	35.432

TABLE A.3(a): Efficiency (p = 2)

[* Denotes maximum in each column]

λ	r = 1	r = 2	r = 4	r = 8
0.0	.375	.500	.600	.667
0.1	.392	.520	.621	.688
0.2	.406	.536	.638	.705
0.3	.418	.549	.651	.718
0.4	.428	.560	.662	.728
0.5	.436	.569	.671	.737
0.6	.444	.576	.678	.743
0.7	.450	.583	.684	.749
0.8	.455	.588	.688	.753
0.9	.460	.592	.692	.756
1.0	.464	.596	.695	.758
1.1	.467	.599	.697	.760
1.2	.470	.601	.699	.761
1.3	.472	.603	.700	.761*
1.4	.474	.604	.700	.761
1.5	.476	.605	.700*	.760
1.6	.477	.606	.700	.759
1.7	.478	.606*	.699	.758
1.8	.478	.606	.698	.756
1.9	.479	.605	.697	.755
2.0	.479*	.604	.696	.753
2.5	.476	.597	.685	.739
3.0	.469	.587	.670	.722
3.5	.460	.573	.653	.703
4.0	.449	.558	.635	.682
4.5	.437	.542	.615	.660
5.0	.425	.525	.595	.637

TABLE A.3(b): Efficiency (p = 3)

[* denotes maximum in each column]

λ	r = 1	r = 2	r = 4	r = 8
0.0	.405	.540	.648	.721
0.1	.415	.552	.660	.732
0.2	.424	.561	.669	.741
0.3	.431	.569	.677	.748
0.4	.438	.576	.684	.754
0.5	.443	.581	.689	.759
0.6	.448	.586	.693	.763
0.7	.452	.590	.697	.766
0.8	.456	.594	.700	.768
0.9	.459	.597	.702	.770
1.0	.462	.599	.704	.771
1.1	.464	.601	.705	.772
1.2	.466	.603	.706	.772*
1.3	.468	.604	.707	.772
1.4	.469	.605	.707*	.772
1.5	.471	.605	.707	.771
1.6	.472	.606	.706	.770
1.7	.472	.606*	.706	.769
1.8	.473	.606	.705	.768
1.9	.473	.605	.704	.766
2.0	.473*	.605	.702	.764
2.5	.472	.600	.694	.753
3.0	.467	.591	.682	.739
3.5	.461	.581	.668	.723
4.0	.452	.569	.653	.705
4.5	.443	.556	.637	.687
5.0	.434	.542	.620	.668

TABLE A.3(c): Efficiency (p = 4)

[* denotes maximum in each column]

λ	r = 1	r = 2	r = 4	r = 8
0.0	.422	.562	.675	.750
0.1	.428	.570	.682	.757
0.2	.434	.576	.688	.762
0.3	.439	.581	.693	.767
0.4	.444	.586	.697	.770
0.5	.448	.589	.700	.773
0.6	.451	.593	.703	.775
0.7	.454	.596	.705	.777
0.8	.457	.598	.707	.778
0.9	.459	.600	.709	.779
1.0	.461	.602	.710	.780
1.1	.463	.603	.711	.780*
1.2	.465	.604	.711	.780
1.3	.466	.605	.711*	.780
1.4	.467	.606	.711	.779
1.5	.468	.606	.711	.778
1.6	.469	.606	.711	.777
1.7	.469	.606*	.710	.776
1.8	.470	.606	.709	.775
1.9	.470	.606	.708	.773
2.0	.470*	.605	.707	.772
2.5	.469	.601	.700	.762
3.0	.466	.594	.690	.750
3.5	.461	.586	.678	.736
4.0	.454	.576	.665	.720
4.5	.447	.565	.651	.704
5.0	.439	.553	.636	.688

TABLE A.3(d): Efficiency (p = 5)

[* denotes maximum in each column]

λ	r = 1	r = 2	r = 4	r = 8
0.0	.432	.576	.692	.769
0.1	.437	.581	.696	.773
0.2	.441	.586	.700	.776
0.3	.445	.589	.703	.779
0.4	.448	.592	.706	.781
0.5	.451	.595	.708	.783
0.6	.454	.598	.710	.784
0.7	.456	.600	.712	.785
0.8	.458	.601	.713	.786
0.9	.460	.603	.714	.786
1.0	.462	.604	.715	.786*
1.1	.463	.605	.715	.786
1.2	.464	.606	.715	.786
1.3	.465	.607	.715*	.785
1.4	.466	.607	.715	.785
1.5	.467	.607	.715	.784
1.6	.467	.607*	.714	.783
1.7	.468	.607	.713	.782
1.8	.468	.607	.713	.781
1.9	.468	.607	.712	.779
2.0	.469*	.606	.711	.778
2.5	.468	.603	.704	.769
3.0	.465	.597	.695	.758
3.5	.461	.589	.685	.745
4.0	.456	.581	.673	.731
4.5	.449	.571	.661	.717
5.0	.443	.561	.648	.702

TABLE A.3(e): Efficiency (p = 6)

[* denotes maximum in each column]

λ	r = 1	r = 2	r = 4	r = 8
0.0	.439	.586	.703	.781
0.1	.443	.589	.706	.784
0.2	.446	.592	.709	.786
0.3	.449	.595	.711	.788
0.4	.451	.598	.713	.789
0.5	.454	.600	.714	.790
0.6	.456	.601	.716	.791
0.7	.458	.603	.717	.791
0.8	.459	.604	.718	.792
0.9	.461	.605	.718	.792*
1.0	.462	.606	.718	.792
1.1	.463	.607	.719	.791
1.2	.464	.608	.719*	.791
1.3	.465	.608	.718	.790
1.4	.466	.608	.718	.789
1.5	.466	.608	.718	.789
1.6	.467	.608*	.717	.788
1.7	.467	.608	.716	.786
1.8	.467	.608	.716	.785
1.9	.468	.608	.715	.784
2.0	.468*	.607	.714	.782
2.5	.467	.604	.708	.774
3.0	.465	.599	.700	.764
3.5	.461	.592	.690	.753
4.0	.457	.584	.680	.740
4.5	.451	.576	.668	.727
5.0	.445	.567	.657	.713

TABLE A.3(f): Efficiency (p = 7)

[* denotes maximum in each column]

λ	r = 1	r = 2	r = 4	r = 8
0.0	.445	.593	.711	.790
0.1	.447	.595	.714	.792
0.2	.450	.598	.715	.794
0.3	.452	.600	.717	.795
0.4	.454	.602	.718	.795
0.5	.456	.603	.719	.796
0.6	.457	.604	.720	.796
0.7	.459	.606	.721	.797*
0.8	.460	.607	.721	.797
0.9	.461	.607	.722	.796
1.0	.462	.608	.722	.796
1.1	.463	.609	.722*	.796
1.2	.464	.609	.722	.795
1.3	.465	.609	.721	.794
1.4	.466	.609	.721	.793
1.5	.466	.610*	.720	.793
1.6	.466	.609	.720	.792
1.7	.467	.609	.719	.790
1.8	.467	.609	.718	.789
1.9	.467	.609	.718	.788
2.0	.467*	.608	.717	.787
2.5	.467	.605	.711	.779
3.0	.465	.600	.703	.769
3.5	.461	.595	.695	.759
4.0	.457	.588	.685	.747
4.5	.453	.580	.675	.735
5.0	.447	.572	.664	.722

TABLE A.3(g): Efficiency (p = 8)

[* denotes maximum in each column]

λ	r = 1	r = 2	r = 4	r = 8
0.0	.449	.598	.718	.798
0.1	.451	.600	.719	.799
0.2	.453	.602	.721	.799
0.3	.455	.603	.722	.800
0.4	.456	.605	.723	.800
0.5	.458	.606	.723	.801
0.6	.459	.607	.724	.801*
0.7	.460	.608	.724	.801
0.8	.461	.609	.724	.801
0.9	.462	.609	.725*	.800
1.0	.463	.610	.725	.800
1.1	.464	.610	.724	.799
1.2	.465	.610	.724	.799
1.3	.465	.611	.724	.798
1.4	.466	.611*	.723	.797
1.5	.466	.611	.723	.796
1.6	.466	.611	.722	.795
1.7	.467	.610	.722	.794
1.8	.467	.610	.721	.793
1.9	.467	.610	.720	.791
2.0	.467*	.609	.719	.790
2.5	.466	.606	.713	.783
3.0	.465	.602	.707	.774
3.5	.462	.597	.699	.764
4.0	.458	.590	.690	.753
4.5	.454	.583	.680	.742
5.0	.449	.575	.670	.730

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

18 ARO 19 11959.19-m

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 11959.19-m	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) 6 Some Statistical Procedures Based on Distances.	5. TYPE OF REPORT & PERIOD COVERED 9 TECHNICAL rept.
	6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1096
7. AUTHOR(s) 10 Joseph J./Walker	8. CONTRACT OR GRANT NUMBER(s) 15 DAAG29-74-C-0030, NSF-CP-42325 14 MMS-1096
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina Chapel Hill, North Carolina 27514	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 12 11p.
11. CONTROLLING OFFICE NAME AND ADDRESS Army Research Office Box 1221 Research Triangle Park, NC 27709	12. REPORT DATE 11 Nov[redacted]76
	13. NUMBER OF PAGES 107
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Distances, Classification, Measurement errors, Asymptotic experiment, Noncentral chi-square distribution, Clustering.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The distributions of statistics employed in classifying the source of a new observation, using observed distances which ~~we~~ are subjected to measurement errors, are discussed. A useful approximate expansion is obtained.

A new method of estimating the parameter of a noncentral chi-square distribution is derived.

182 850

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)